Polar Lattices for Lossy Compression

Ling Liu and Cong Ling *Member, IEEE*

Abstract

Polar lattices, which are constructed from polar codes, have recently been proved to be able to achieve the capacity of the additive white Gaussian noise (AWGN) channel. In this work, we propose a new construction of polar lattices to solve the dual problem, i.e., achieving the rate-distortion bound of a memoryless Gaussian source, which means that polar lattices can also be good for the lossy compression of continuous sources. The structure of the proposed polar lattices enables us to integrate the post-entropy coding process into the lattice quantizer, which simplifies the quantization process. The overall complexity of encoding and decoding complexity is $O(N \log^2 N)$ for a sub-exponentially decaying excess distortion. Moreover, the nesting structure of polar lattices further provides solutions for some multi-terminal coding problems. The Wyner-Ziv coding problem for a Gaussian source can be solved by an AWGN capacity-achieving polar lattice nested in a rate-distortion bound achieving one, and the Gelfand-Pinsker problem can be solved in a reversed manner.

I. Introduction

Vector quantization (VQ) [1] has been widely used for source coding of image and speech data since the 1980s. Compared with scalar quantization, the advantage of VQ, guaranteed by Shannon's rate-distortion theory, is that better performance can always be achieved by coding vectors instead of scalars, even in the case of memoryless sources. However, the Shannon theory does not provide us any constructive VQ design scheme. During the past several decades, many practical VQ techniques with relatively low complexity have been proposed, such as lattice VQ [2], multistage VQ [3], tree-structured VQ [4], gain-shape VQ [5], etc. Among them, lattice VQ is of particular interest because its highly regular structure makes compact storage and fast quantization possible.

In this work, we present an explicit construction of polar lattices for quantization, which achieves the rate-distortion bound of the continuous Gaussian source. It is well known that the optimal output alphabet size is infinite for continuous-amplitude sources. Particularly, the rate distortion function for the Gaussian source of variance σ_s^2 under the squared-error distortion measure $d(x, y) = \|x - y\|^2$ is given by

$$R(\Delta) = \max\left\{\frac{1}{2}\log\left(\frac{\sigma_s^2}{\Delta}\right), 0\right\}, \tag{1}$$

where Δ and R denote the average distortion and rate per symbol, respectively. However, in practice, the size of the reconstruction alphabet needs to be finite. Unconstructively, [6, Theorem 9.6.2] shows the existence of

This work was presented in part at the IEEE Inform. Theory Workshop (ITW) 2015, Jeju Island, Korea, October, 2015. This work was supported in part by the China Scholarship Council.

Ling Liu and Cong Ling are with the Department of Electrical and Electronic Engineering, Imperial College London, London, UK (e-mails: l.liu12@imperial.ac.uk, cling@ieee.org).

a block code with finite number of output letters that achieves performance arbitrarily close to the rate-distortion bound. Then a size-constrained output alphabet rate-distortion function $R_M(\Delta)$ was defined in [7] with M denoting the size of output alphabet. The well-known trellis coded quantization (TCQ) [8] was motivated by this alphabet constrained rate-distortion theory. It was shown that for a given encoding rate of R bits per symbol, the rate-distortion function $R(\Delta)$ can be approached by using a TCQ encoder with rate $R+1$ after an initial Lloyd-Max quantization. It is equivalent to the trellis coded modulation (TCM) in the sense that m information bits are transmitted using 2^{m+1} constellation points. A near-optimum lattice quantization scheme based on tailbiting convolutional codes was introduced in [9]. Despite enjoying a good practical performance, a theoretical proof of the rate-distortion bound achieving TCQ with low complexity is still missing. More recently, a scheme based on low density Construction-A (LDA) lattices [10] was proved to be quantization-good (defined in Sect. III) using the minimum-distance lattice decoder. However, in practice the ideal performance cannot be realized by the suboptimal belief-propagation decoding algorithm.

Polar lattices have the potential in solving this problem with low complexity. As shown in [11], this class of lattices allows us to employ the discrete Gaussian distribution for lattice shaping. This distribution shares many similar properties to the continuous Gaussian distribution and obtains the optimal shaping gain when its associated flatness factor is negligible. We may use the discrete Gaussian distribution instead of the continuous one as the distribution of the reconstruction alphabet. This idea has already been proposed in [12] for random lattice quantization. It is also shown in [11] that even using binary lattice partition, the number of the partition levels r does not need to be very large ($O(\log\log N)$) to achieve the capacity $\frac{1}{2}\log(1+\mathsf{SNR})$ of the additive white Gaussian noise (AWGN) channel, where SNR denotes the signal noise ratio. By the duality between source coding and channel coding, the quantization lattices can be roughly viewed as a channel coding lattice constructed on the test channel. For a Gaussian source with variance σ_s^2 and an average distortion Δ, the test channel is actually an AWGN channel with noise variance Δ. In this case, the "SNR" of the test channel is $\frac{\sigma_s^2-\Delta}{\Delta}$, and its "capacity" is $\frac{1}{2}\log(\frac{\sigma_s^2}{\Delta})$, which implies that the rate of the polar lattice quantizer can be made arbitrarily close to $\frac{1}{2}\log(\frac{\sigma_s^2}{\Delta})$. Therefore, based on this idea, we propose the construction of polar lattices which are good for quantization in this work. We note that the difference between the quantization polar lattices and the AWGN channel coding polar lattices not only lies in the construction of their component polar codes, but also in the role of their associate flatness factors. For the AWGN channel coding polar lattices, the flatness factor is required to be negligible to ensure a coding rate close to the AWGN capacity and it has no influence on the error correction performance. For the quantization polar lattices, however, the flatness factor affects both the compression rate and the distortion performance. This is also the reason why the lattice Gaussian distribution can be optimal for both channel coding and quantization simultaneously (see Remark 1), and consequently be utilized for Gaussian Wyner-Ziv and Gelfand-Pinsker coding.

A. Our contribution

The novel technical contribution of this paper is two-fold:

- The construction of polar lattices for the Gaussian source and the proof of their rate-distortion bound achieving.

This is a dual work of capacity-achieving polar lattices for the AWGN channel, and it can also be considered as an extension of binary polar lossy coding to the multilevel coding scenario. Compared with traditional lattice quantization schemes [13], [14], which generally require a separate entropy encoding process after obtaining the quantized lattice points, our scheme naturally integrates these two processes together. The analysis of these quantization polar lattices prepares us for the further discussion of Gaussian Wyner-Ziv and Gelfand-Pinsker problems.

- The solutions of the Gaussian Wyner-Ziv and Gelfand-Pinsker problems, which consist of two nested polar lattices. One is AWGN capacity-achieving and the other is Gaussian rate-distortion bound achieving. The two lattices are simultaneously shaped according to a proper lattice Gaussian distribution. Note that the Wyner-Ziv and Gelfand-Pinsker problems for the binary case have been solved by Korada and Urbanke [15] using nested polar codes. However, in the Gaussian case, the problems turn out to be more complicated as the Wyner-Ziv bound becomes lower (the Gelfand-Pinsker capacity becomes larger by duality). As mentioned in [16], the extremely severe conditions [16, eq. (12)] and [16, eq. (18)] for the bound, which corresponds to the scenario where both encoder and decoder know the side information, can be satisfied in the Gaussian case rather than the binary case because of infinite alphabet size, meaning that more effort should be made for the Gaussian case. As a result, our polar lattice coding scheme achieves the whole region of the Wyner-Ziv bound and has no requirement on the signal noise ratio for the Gelfand-Pinsker capacity.

B. Relation to Prior Works

As mentioned above, although the TCQ technique performs well in practice, its theoretical limit is still unclear, to the best of our knowledge. Polar lattices, as we will see, can be theoretically proved to be able to achieve the rate-distortion bound. Moreover, thanks to their low complexity, considerably high-dimensional polar lattices are available in practice, providing a quantization performance with gap less than 0.2 dB to the achievable bound when the lattice dimension $N = 2^{18}$.

The sparse regression codes were also proved to achieve the the optimal rate-distortion bound of i.i.d Gaussian sources with polynomial complexity [17], [18]. In fact, there exists a trade-off between the distortion performance and encoding complexity. For a block length N, typical encoding complexity of this kind of codes is $O((N/\log N)^2)$ for an exponentially decaying excess distortion with exponent $O(N/\log N)$, and their designed random matrix incurs $N \times O(N^2)$ storage complexity. In comparison, the construction of polar lattices is as explicit as that of polar codes themselves, and the complexity is quasi-linear $O(N \log^2 N)$ for a sub-exponentially decaying excess distortion with exponent roughly $O(\sqrt{N})$.

The saliently nesting structure of polar lattices also gives us solutions to the Gaussian Wyner-Ziv and Gelfand-Pinsker problems. According to the prior work by Zamir, Shamai and Erez [19], [20], the two problems can be solved by nested quantization-good and AWGN-good lattices. However, due to the lack of explicit construction of such good lattices, no explicit solution was addressed. A practical scheme based on multidimensional nested lattice codes for the Gaussian Wyner-Ziv problem was also proposed in [21]. The performance of this scheme can be

very close to the Wyner-Ziv bound but a theoretical proof is still missing. A lattice-based Gelfand-Pinsker coding scheme using repeat-accumulate codes, which were concatenated with trellis shaping, was also presented in [22]. This scheme was shown to be able to obtain a very close-to-capacity performance. Unfortunately, the complexity grows exponentially to achieve the shaping gain and a theoretical proof for the Gelfand-Pinsker capacity-achieving is also missing. In this work, we solve these problems by combining the recently proposed AWGN capacity achieving polar lattices [11] and the rate-distortion bound achieving ones.

C. Outline of the paper

The paper is organized as follows: Section II presents the background of polar codes and polar lattices. The construction of rate-distortion bound achieving polar lattices is investigated in Section III. Some simulation results of polar lattices for quantization are given in Section IV. In Section V and Section VI, the solutions of the Gaussian Wyner-ziv and Gelfand-Pinsker problems are addressed accordingly, by combining the AWGN capacity achieving polar lattices and the proposed quantization polar lattices. The paper is concluded in Section VII.

D. Notations

All random variables (RVs) will be denoted by capital letters. Let P_X denote the probability distribution of a RV X taking values x in a set $\mathcal{X}$. For multilevel coding, we denote by X_ℓ a RV X at level ℓ. The i-th realization of X_ℓ is denoted by x_ℓ^i. We also use the notation $x_\ell^{i:j}$ as a shorthand for a vector $(x_\ell^i, ..., x_\ell^j)$, which is a realization of RVs $X_\ell^{i:j} = (X_\ell^i, ..., X_\ell^j)$. Similarly, $x_{\ell:\jmath}^i$ will denote the realization of the i-th RV from level ℓ to level $\jmath$, i.e., of $X_{\ell:\jmath}^i = (X_\ell^i, ..., X_\jmath^i)$. For a set $\mathcal{I}$, $|\mathcal{I}|$ represents its cardinality. For an integer N, $[N]$ denotes the set of all integers from 1 to N. $\mathbb{1}(\cdot)$ denotes an indicator function. Let $I(X;Y)$ denote the mutual information between X and Y. The notations $R \to I(X;Y)^+$ and $R \to I(X;Y)^-$ will be used to represent a rate approaching $I(X;Y)$ from the right side (equal or greater than $I(X;Y)$) and the left side (equal or less than $I(X;Y)$), respectively. Throughout this paper, we use the binary logarithm and information is measured in bits.

II. Background

A. Lattice Codes and lattice Gaussian distribution

A lattice is a discrete subgroup of $\mathbb{R}^n$ which can be described by

$$\Lambda = \{\lambda = \mathbf{B}z : z \in \mathbb{Z}^n\}, \tag{2}$$

where the columns of the generator matrix $\mathbf{B} = [\mathrm{b}_1, \cdots, \mathrm{b}_n]$ are assumed to be linearly independent.

For a vector $x \in \mathbb{R}^n$, the nearest-neighbor quantizer associated with Λ is $Q_\Lambda(x) = \arg\min_{\lambda\in\Lambda} \|\lambda - x\|$, where ties are resolved arbitrarily. We define the modulo lattice operation by $x \bmod \Lambda \triangleq x - Q_\Lambda(x)$. The Voronoi region of Λ, defined by $\mathcal{V}(\Lambda) = \{x : Q_\Lambda(x) = 0\}$, specifies the nearest-neighbor decoding region. The volume of a fundamental region is equal to that of the Voronoi region $\mathcal{V}(\Lambda)$, which is given by $V(\Lambda) = |\det(\mathbf{B})|$.

For $\sigma > 0$ and $c \in \mathbb{R}^n$, we define the Gaussian distribution of variance σ^2 centered at c as

$$f_{\sigma,c}(x) = \frac{1}{(\sqrt{2\pi}\sigma)^n} e^{-\frac{\|x-c\|^2}{2\sigma^2}}, \ x \in \mathbb{R}^n.$$

Let $f_{\sigma,0}(x) = f_\sigma(x)$ for short. The Λ-periodic function is defined as

$$f_{\sigma,\Lambda}(x) = \sum_{\lambda\in\Lambda} f_{\sigma,\lambda}(x) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \sum_{\lambda\in\Lambda} e^{-\frac{\|x-\lambda\|^2}{2\sigma^2}}.$$

We note that $f_{\sigma,\Lambda}(x)$ is a probability density function (PDF) if x is restricted to the fundamental region $\mathcal{R}(\Lambda)$. This distribution is actually the PDF of the Λ-aliased Gaussian noise, i.e., the Gaussian noise after the mod-Λ operation [23].

The flatness factor of a lattice Λ is defined as [24]

$$\epsilon_\Lambda(\sigma) \triangleq \max_{x\in\mathcal{R}(\Lambda)} |V(\Lambda) f_{\sigma,\Lambda}(x) - 1|.$$

We define the discrete Gaussian distribution over Λ centered at c as the discrete distribution taking values in $\lambda \in \Lambda$:

$$D_{\Lambda,\sigma,c}(\lambda) = \frac{f_{\sigma,c}(\lambda)}{f_{\sigma,c}(\Lambda)}, \ \forall \lambda \in \Lambda, \tag{3}$$

where $f_{\sigma,c}(\Lambda) = \sum_{\lambda\in\Lambda} f_{\sigma,c}(\lambda)$. For convenience, we write $D_{\Lambda,\sigma} = D_{\Lambda,\sigma,0}$. It has been proved to achieve the optimum shaping gain when the flatness factor is negligible [25].

A sublattice $\Lambda' \subset \Lambda$ induces a partition (denoted by Λ/Λ') of Λ into equivalence groups modulo Λ'. The order of the partition is denoted by $|\Lambda/\Lambda'|$, which is equal to the number of the cosets. If $|\Lambda/\Lambda'| = 2$, we call this a binary partition. Let $\Lambda(\Lambda_0)/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'(\Lambda_r)$ for $r \geq 1$ be an n-dimensional lattice partition chain. If only one level is applied ($r = 1$), the construction is known as "Construction A". If multiple levels are used, the construction is known as "Construction D" [2, p.232]. For each partition $\Lambda_{\ell-1}/\Lambda_\ell$ ($1 \leq \ell \leq r$) a code C_ℓ over $\Lambda_{\ell-1}/\Lambda_\ell$ selects a sequence of coset representatives a_ℓ in a set A_ℓ of representatives for the cosets of Λ_ℓ. This construction requires a set of nested linear binary codes C_ℓ with block length N and dimension of information bits k_ℓ which are represented as $[N, k_\ell]$ for $1 \leq \ell \leq r$ and $C_1 \subseteq C_2 \cdots \subseteq C_r$. Let ψ be the natural embedding of $\mathbb{F}_2^N$ into $\mathbb{Z}^N$, where $\mathbb{F}_2$ is the binary field. Consider $\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_N$ be a basis of $\mathbb{F}_2^N$ such that $\mathbf{g}_1, \cdots \mathbf{g}_{k_\ell}$ span C_ℓ. When $n = 1$, the binary lattice L consists of all vectors of the form

$$\sum_{\ell=1}^{r} 2^{\ell-1} \sum_{j=1}^{k_\ell} u_\ell^j \psi(\mathbf{g}_j) + 2^r z, \tag{4}$$

where $u_\ell^j \in \{0, 1\}$ and $z \in \mathbb{Z}^N$.

B. Polar codes and polar lattices

Polar codes are the first kind of codes that can be proved to be able to achieve the capacity of any binary memoryless symmetric (BMS) channel. Let $\tilde{W}(y|x)$ be a BMS channel with input alphabet $\mathcal{X} = \{0, 1\}$ and output alphabet $\mathcal{Y} \subseteq \mathbb{R}$. Given the capacity $I(\tilde{W})$ of $\tilde{W}$ and any rate $R < I(\tilde{W})$, the information bits of a polar code with block length $N = 2^m$ are indexed by a set of RN rows of the generator matrix $G_N = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 1 \end{smallmatrix}\right]^{\otimes m}$, where $\otimes$

denotes the Kronecker product. The matrix G_N combines N identical copies of $\tilde{W}$ to $\tilde{W}_N$. Then this combination can be successively split into N BMS subchannels, denoted by $\tilde{W}_N^{(i)}$ with $1 \leq i \leq N$. By channel polarization, the fraction of good (roughly error-free) subchannels is about $I(\tilde{W})$ as $m \to \infty$. Therefore, to achieve capacity, information bits should be sent over those good subchannels and the rest are fed with frozen bits which are known before transmission. The indices of good subchannels can be identified according their associate Bhattacharyya Parameters.

Definition 1 (Bhattacharyya Parameter for Symmetric Channel [26]): Given a BMS channel $\tilde{W}$ with transition probability $P_{Y|X}$, the Bhattacharyya parameter $\tilde{Z} \in [0,1]$ is defined as

$$\tilde{Z}(\tilde{W}) \quad \triangleq \sum_y \sqrt{P_{Y|X}(y|0)P_{Y|X}(y|1)}.$$

Based on the Bhattacharyya parameter, the information set $\tilde{\mathcal{I}}$ is defined as $\{i : \tilde{Z}(\tilde{W}_N^{(i)}) \leq 2^{-N^\beta}\}$, and the frozen set $\tilde{\mathcal{F}}$ is the complement of $\tilde{\mathcal{I}}$. Let P_B denote the block error probability of a polar code under SC decoding. It can be upper-bounded as $P_B \leq \Sigma_{i\in\tilde{\mathcal{I}}}\tilde{Z}(\tilde{W}_N^{(i)})$. An efficient algorithm to evaluate the Bhattacharyya parameter of subchannels when $\tilde{W}$ is a binary $\Lambda_{\ell-1}/\Lambda_\ell$ channel has been presented in [11], based on the prior work on the construction of polar codes [27], [28].

A polar lattice is constructed by using a set of nested polar codes as the component codes in (4). An explicit construction of AWGN-good polar lattices (defined in [11]) based on the multilevel approach of Forney *et al.* [23] has been presented in [11]. The key idea is to design a polar code to achieve the capacity for each level $\ell = 1, 2, ..., r$ in Construction D.

To achieve the capacity of the power-constrained Gaussian channel, we need to apply Gaussian shaping over the AWGN-good polar lattice, which is considered to be difficult to do directly. Motivated by [25], we can apply Gaussian shaping to the top lattice Λ instead. This shaping process generally leads to a nonuniform input distribution and an binary memoryless asymmetric (BMA) channel for each partition. In this case, we need the recently introduced polar codes for asymmetric channels.

Definition 2 (Bhattacharyya Parameter for BMA Channel [29], [30]): Let W be a BMA channel with input $X \in \mathcal{X} = \{0,1\}$ and output $Y \in \mathcal{Y}$, and let P_X and $P_{Y|X}$ denote the input distribution and channel transition probability, respectively. The Bhattacharyya parameter Z for channel W is the defined as

$$\begin{aligned} Z(X|Y) &= 2\sum_y P_Y(y)\sqrt{P_{X|Y}(0|y)P_{X|Y}(1|y)} \\ &= 2\sum_y \sqrt{P_{X,Y}(0,y)P_{X,Y}(1,y)}. \end{aligned}$$

Note that Definition 2 is the same as Definition 1 when P_X is uniform.

Let $X^{1:N}$ and $Y^{1:N}$ be the input and output vector after N independent uses of W. We have the following

property of the polarized random variables $U^{1:N} = X^{1:N}G_N$:

$$\begin{cases} \lim\limits_{N\to\infty} \dfrac{1}{N} \left| \left\{ i : Z(U^i|U^{1:i-1}, Y^{1:N}) \leq 2^{-N^\beta} \text{ and } Z(U^i|U^{1:i-1}) \geq 1 - 2^{-N^\beta} \right\} \right| = I(X;Y), \\ \lim\limits_{N\to\infty} \dfrac{1}{N} \left| \left\{ i : Z(U^i|U^{1:i-1}, Y^{1:N}) \geq 2^{-N^\beta} \text{ or } Z(U^i|U^{1:i-1}) \leq 1 - 2^{-N^\beta} \right\} \right| = 1 - I(X;Y), \end{cases} \tag{5}$$

which provides us a method of achieving the capacity of a BMA channel. Moreover, The Bhattacharyya parameter of a BMA channel can be related to that of a BMS channel, and the decoding of a polar code for the BMA channel can also be converted to that for the BMS channel. (See [11] for more details.)

III. Polar Lattices for Quantization

Let $Y \sim N(0, \sigma_s^2)$ denote a one dimensional Gaussian source with zero mean and variance σ_s^2. Let $Y^{1:N}$ ($\underline{Y}$) be N independent copies of Y and $y^{1:N}$ ($\underline{y}$) be a realization of $Y^{1:N}$. The PDF of $\underline{Y}$ is given by $f_{\underline{Y}}(\underline{y}) = f_{\sigma_s}(\underline{y})$. For an N-dimensional polar lattice L and its associated quantizer $Q_L(\cdot)$, the average distortion Δ after quantization is given by

$$\Delta = \frac{1}{N} \int_{\mathbb{R}^N} \| \underline{y} - Q_L(\underline{y}) \|^2 \, f_{\underline{Y}}(\underline{y}) d\underline{y}. \tag{6}$$

The normalized second moment (NSM) of a quantization lattice L is defined as

$$G(L) = \frac{\frac{1}{N} \int_{\mathcal{V}(L)} \|v\|^2 dv}{V(L)^{1+2/N}}, \tag{7}$$

where vector v is uniformly distributed in $\mathcal{V}(L)$.

Definition 3: An N-dimensional lattice L is called quantization-good [13] if

$$\lim_{N\to\infty} G(L) = \frac{1}{2\pi e}. \tag{8}$$

In [14], an entropy-coded dithered quantization (ECDQ) scheme based on quantization-good lattices was proposed to achieve the rate-distortion bound (1). This scheme requires a pre-shared dither which is uniformly distributed in the Voronoi region of a quantization-good lattice and an entropy encoder after lattice quantization. For our quantization scheme, we will show that dither is not necessary and the entropy encoder can be integrated in the lattice quantization process, which brings much convenience for practical application.

Our task is to construct a polar lattice which achieves the rate distortion bound of the Gaussian source with reconstruction distribution $D_{\Lambda,\sqrt{\sigma_s^2-\Delta}}$. Following the notation of AWGN-good polar lattices [11], we use X to denote the reconstruction alphabet. Firstly, we prove that the rate achieved by $D_{\Lambda,\sqrt{\sigma_s^2-\Delta}}$ can be arbitrarily close to $\frac{1}{2}\log(\frac{\sigma_s^2}{\Delta})$. Note that the following theorem is essentially the same as Theorem 2 in [25]. Here we just reexpress it in the source coding formulation.

Theorem 1 ([25]): Consider a Gaussian test channel where the reconstruction constellation X has a discrete Gaussian distribution $D_{\Lambda-\mathrm{c},\sigma_r}$ for arbitrary $\mathrm{c} \in \mathbb{R}^n$, and where $\sigma_r^2 = \sigma_s^2 - \Delta$ with Δ being the average distortion.

Let $\tilde{\sigma}_\Delta \triangleq \frac{\sigma_r\sqrt{\Delta}}{\sigma_s}$. Then, if $\epsilon = \epsilon_\Lambda(\tilde{\sigma}_\Delta) < \frac{1}{2}$ and $\frac{\pi\epsilon_t}{1-\epsilon_t} < \epsilon$ where

$$\epsilon_t \triangleq \begin{cases} \epsilon_\Lambda(\sigma_r/\sqrt{\frac{\pi}{\pi-t}}), & t \geq 1/e \\ (t^{-4}+1)\epsilon_\Lambda(\sigma_r/\sqrt{\frac{\pi}{\pi-t}}), & 0 < t < 1/e \end{cases} \tag{9}$$

the discrete Gaussian constellation results in mutual information $I_\Delta \geq \frac{1}{2}\log(\frac{\sigma_s^2}{\Delta}) - \frac{5\epsilon}{n}$ per channel use.

The statement of Theorem 1 is non-asymptotical, i.e., it can hold even if $n = 1$. Therefore, it is possible to construct a good polar lattice over one-dimensional lattice partition such as $\mathbb{Z}/2\mathbb{Z}/4\mathbb{Z}$.... The flatness factor ϵ can be made negligible by scaling this binary partition. This technique has already been used to construct AWGN capacity-achieving polar lattices [11].

Note that when the test channel is chosen to be an AWGN channel with noise variance Δ and the reconstruction alphabet is discrete Gaussian distributed, the source distribution is not exactly a continuous Gaussian distribution. In fact, it is a distribution obtained by adding a continuous Gaussian of variance Δ to a discrete Gaussian $D_{\Lambda-\mathrm{c},\sigma_r}$, which is expressed as the following convolution

$$f_{Y'}(y') = \frac{1}{f_{\sigma_r}(\Lambda - c)} \sum_{\lambda\in\Lambda-c} f_{\sigma_r}(\lambda) f_\sigma(y'-\lambda), y' \in \mathbb{R}^n, \tag{10}$$

where $\sigma = \sqrt{\Delta}$ and Y' denotes the new source. For simplicity, in this work we only consider one dimensional binary partition chain ($n = 1$) and Y' is also a one dimensional source.

Therefore, we are actually quantizing source Y' instead of Y using the discrete Gaussian distribution. However, when the flatness factor $\epsilon_\Lambda(\tilde{\sigma}_\Delta)$ is small, a good quantizer constructed from polar lattices for the source Y' is also good for source Y because of the following lemma. The relationship between the quantization of source Y' and Y is shown in Fig. 1.

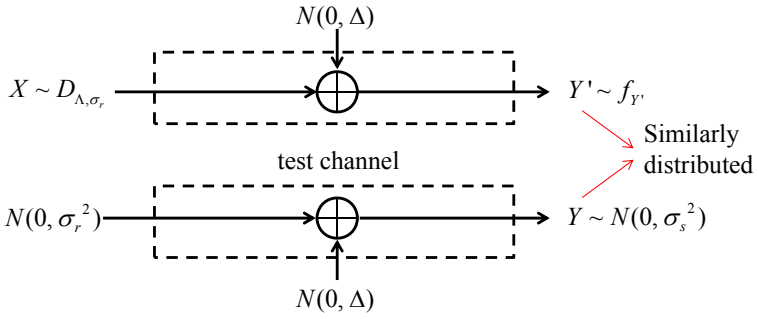


Fig. 1. The relationship between the quantization of source Y' and Y.

Lemma 1 ([25]): If $\epsilon = \epsilon_\Lambda(\tilde{\sigma}_\Delta) < \frac{1}{2}$, the variational distance between the density $f_{Y'}$ of source Y' defined in (10) and the Gaussian density f_Y satisfies $\mathbb{V}(f_{Y'}, f_Y) \leq 4\epsilon$.

Now we consider the construction of polar lattices for quantization. Firstly, we consider the quantization of source Y' using the reconstruction distribution D_{Λ,σ_r}. Since binary partition is used, X can be represented by a binary

string $X_{1:r}$, and we have $\lim_{r\to\infty} P_{X_{1:r}} = P_X = D_{\Lambda,\sigma_r}$. Because the polar lattices are constructed by "Construction D", we are interested in the test channel at each level. Similar to the setting of shaping for AWGN-good polar lattices [11], given the previous $x_{1:\ell-1}$ and the coset $\mathcal{A}_\ell$ determined by $x_{1:\ell}$, the channel transition PDF at level ℓ is

$$\begin{aligned} &P_{Y'|X_\ell,X_{1:\ell-1}}(y'|x_\ell,x_{1:\ell-1}) \\ &= \frac{\sum_{a\in\mathcal{A}_\ell(x_{1:\ell})} P(a)P_{Y'|A}(y'|a)}{P\{\mathcal{A}_\ell(x_{1:\ell})\}} \\ &= \exp\left(-\frac{y'^2}{2(\sigma_s^2+\Delta)}\right)\frac{1}{f_{\sigma_s}(\mathcal{A}_\ell(x_{1:\ell}))}\frac{1}{2\pi\sqrt{\Delta}\sigma_s}\sum_{a\in\mathcal{A}_\ell(x_{1:\ell})}\exp\left(-\frac{1}{2\tilde{\sigma}_\Delta^2}\left(|\alpha y'-a|^2\right)\right), \end{aligned} \tag{11}$$

where $\alpha = \frac{\sigma_r^2}{\sigma_r^2+\Delta}$ is equal to the MMSE coefficient and $\tilde{\sigma}_\Delta = \frac{\sigma_r\sqrt{\Delta}}{\sigma_s}$. Consequently, using D_{Λ,σ_r} as the constellation, the ℓ-th channel is generally asymmetric with the input distribution $P_{X_\ell|X_{1:\ell-1}}$ $(\ell \le r)$, which can be calculated according to the definition of D_{Λ,σ_r}.

The lattice quantization can be viewed as lossy compression for all binary testing channels from level 1 to r. Here we start with the first level. Let $y'^{1:N}$ denotes the realization of N i.i.d copies of source Y'. Although Y' is a continuous source with density given by (10) and y' is drawn from $\mathbb{R}$, to keep the notations consistent (the definition of the Bhattacharyya parameter [26] is given by a summation), from now on we will express the distortion measurement as well as the variantional distance in the form of summation instead of integral.

Since the test channel at each level is not necessarily symmetric and the reconstruction constellation is not uniformly distributed, we have to consider the lossy compression for nonuniform source and asymmetric distortion measure. The solution of this problem has already been introduced in [30], and it turns out to be similar to the construction of capacity-achieving polar codes for asymmetric channels.

For the first level, letting $U_1^{1:N} = X_1^{1:N}G_N$, where G_N is the $N\times N$ generator matrix of polar codes, we define the information set $\mathcal{I}_1$, frozen set $\mathcal{F}_1$ and shaping set $\mathcal{S}_1$ based on the Bhattacharyya parameter as follows:

$$\begin{cases} \mathcal{F}_1 = \{i\in[N] : Z(U_1^i|U_1^{1:i-1},Y'^{1:N}) \ge 1-2^{-N^\beta}\} \\ \mathcal{I}_1 = \{i\in[N] : Z(U_1^i|U_1^{1:i-1}) > 2^{-N^\beta} \text{ and } Z(U_1^i|U_1^{1:i-1},Y'^{1:N}) < 1-2^{-N^\beta}\} \\ \mathcal{S}_1 = \{i\in[N] : Z(U_1^i|U_1^{1:i-1}) \le 2^{-N^\beta}\}. \end{cases} \tag{12}$$

The shaping set $\mathcal{S}_1$ is determined by the distribution P_{X_1}. Note that this definition is similar to that in [11, Equation (21)]. The difference is that $\mathcal{I}_1$ is designed to be slightly larger to guarantee a desired distortion level. The asymmetric Bhattacharyya parameter $Z(U_1^i|U_1^{1:i-1},Y'^{1:N})$ and $Z(U_1^i|U_1^{1:i-1})$ can be efficiently calculated by symmetric Bhattacharyya parameter $\tilde{Z}(\tilde{U}_1^i|\tilde{U}_1^{1:i-1},X_1^{1:N}\oplus\tilde{X}_1^{1:N},Y'^{1:N})$ and $\tilde{Z}(\tilde{U}_1^i|\tilde{U}_1^{1:i-1},X_1^{1:N}\oplus\tilde{X}_1^{1:N})$, respectively (see [11] for more details). According to [11, Theorem 5], the proportion of set $\mathcal{I}_1$ approaches $I(X_1;Y')$ when $N\to\infty$.

After getting $\mathcal{F}_1$, $\mathcal{I}_1$ and $\mathcal{S}_1$, for a source sequence $y'^{1:N}$, the encoder determines $u_1^{1:N}$ according to the following

rule:

$$u_1^i = \begin{cases} 0 \text{ w. p. } P_{U_1^i|U_1^{1:i-1},Y'^{1:N}}(0|u_1^{1:i-1}, y'^{1:N}) \\ 1 \text{ w. p. } P_{U_1^i|U_1^{1:i-1},Y'^{1:N}}(1|u_1^{1:i-1}, y'^{1:N}) \end{cases} \text{ if } i \in \mathcal{I}_1, \tag{13}$$

and

$$u_1^i = \begin{cases} \bar{u}_1^i \quad \text{if} \quad i \in \mathcal{F}_1 \\ \underset{u}{\operatorname{argmax}} P_{U_1^i|U_1^{1:i-1}}(u|u_1^{1:i-1}) \quad \text{if} \quad i \in \mathcal{S}_1. \end{cases} \tag{14}$$

Here $\bar{u}_1^i$ is a uniformly random bit determined before lossy compression. The output of the encoder at level 1 is $u_1^{\mathcal{I}_1} = \{u_1^i, i \in \mathcal{I}_1\}$. To reconstruct $x_1^{1:N}$, the decoder uses the shared $u_1^{\mathcal{F}_1}$ and the received $u_1^{\mathcal{I}_1}$ to recover $u_1^{\mathcal{S}_1}$ according to $\underset{u}{\operatorname{argmax}} P_{U_1^i|U_1^{1:i-1}}(u|u_1^{1:i-1})$ and then $x_1^{1:N} = u_1^{1:N} G_N$. The probability $P_{U_1^i|U_1^{1:i-1}}$ and $P_{U_1^i|U_1^{1:i-1},Y'^{1:N}}$ can both be calculated efficiently by the successive cancellation algorithm with complexity $O(N \log N)$ [11].

Theorem 2: Let $Q_{U_1^{1:N},Y'^{1:N}}(u_1^{1:N}, y'^{1:N})$ denote the joint distribution for $U_1^{1:N}$ and $Y_1^{1:N}$ according to the encoding rule described in (13) and (14). Consider another encoder using the encoding rule (13) for all $i \in [N]$ and let $P_{U_1^{1:N},Y'^{1:N}}(u_1^{1:N}, y'^{1:N})$ denote the resulted joint distribution. For any $\beta' < \beta < 1/2$ satisfying (12) and $R_1 = \frac{|\mathcal{I}_1|}{N} > I(X_1; Y')$,

$$\mathbb{V}(P_{U_1^{1:N},Y'^{1:N}}, Q_{U_1^{1:N},Y'^{1:N}}) = O(2^{-N^{\beta'}}). \tag{15}$$

The same statement has been given in [30] yet without proof. Here we prove the theorem in Appendix A for completeness.

Now we introduce the construction for higher levels. Taking the second level as an example, to make up the reconstruction constellation distribution, the input distribution at level 2 should be $P_{X_2|X_1}$. Based on the quantization results $(U_1^{1:N}, Y'^{1:N})$ given by the encoder at level 1, some U_2^i ($U_2^{1:N} = X_2^{1:N} G_N$) is almost deterministic given $(U_2^{1:i-1}, U_1^{1:N})$. Since there is a one-to-one mapping between $X_1^{1:N}$ and $U_1^{1:N}$, given $(U_2^{1:i-1}, U_1^{1:N})$ is the same as given $(U_2^{1:i-1}, X_1^{1:N})$. We define the information set $\mathcal{I}_2$, frozen set $\mathcal{F}_2$ and shaping set $\mathcal{S}_2$ as follows:

$$\begin{cases} \mathcal{F}_2 = \{i \in [N] : Z(U_2^i|U_2^{1:i-1}, X_1^{1:N}, Y'^{1:N}) \geq 1 - 2^{-N^\beta}\} \\ \mathcal{I}_2 = \{i \in [N] : Z(U_2^i|U_2^{1:i-1}, X_1^{1:N}) > 2^{-N^\beta} \text{ and } Z(U_2^i|U_2^{1:i-1}, X_1^{1:N}, Y'^{1:N}) < 1 - 2^{-N^\beta}\} \\ \mathcal{S}_2 = \{i \in [N] : Z(U_2^i|U_2^{1:i-1}, X_1^{1:N}) \leq 2^{-N^\beta}\}. \end{cases} \tag{16}$$

The proportion of $\mathcal{I}_2$ approaches $I(X_2; Y|X_1)$ when N is sufficiently large [11]. For a given source sequence pair $(u_1^{1:N}, y'^{1:N})$ or $(x_1^{1:N}, y'^{1:N})$, the encoder at level 2 determines $u_2^{1:N}$ according to the following rule:

$$u_2^i = \begin{cases} 0 \text{ w.p. } P_{U_2^i|U_2^{1:i-1},X_1^{1:N},Y'^{1:N}}(0|u_2^{1:i-1}, x_1^{1:N}, y'^{1:N}) \\ 1 \text{ w.p. } P_{U_2^i|U_2^{1:i-1},X_1^{1:N},Y'^{1:N}}(1|u_2^{1:i-1}, x_1^{1:N}, y'^{1:N}) \end{cases} \text{ if } i \in \mathcal{I}_2, \tag{17}$$

and

$$u_2^i = \begin{cases} \bar{u}_2^i \quad \text{if} \quad i \in \mathcal{F}_2 \\ \underset{u}{\operatorname{argmax}} P_{U_2^i|U_2^{1:i-1},X_1^{1:N}}(u|u_2^{1:i-1}, x_1^{1:N}) \quad \text{if} \quad i \in \mathcal{S}_2. \end{cases} \tag{18}$$

We further extend Theorem 2 to the second level.

Theorem 3: Let $Q_{U_2^{1:N},U_1^{1:N},Y'^{1:N}}(u_2^{1:N},u_1^{1:N},y'^{1:N})$ denote the joint distribution for $U_2^{1:N}$ and $(U_1^{1:N},Y_1^{1:N})$ according to the encoding rule described in (17) and (18). Consider another encoder using the encoding rule (17) for all $i\in[N]$ and let $P_{U_2^{1:N},U_1^{1:N},Y'^{1:N}}(u_2^{1:N},u_1^{1:N},y'^{1:N})$ denote the resulted joint distribution. For any $\beta'<\beta<1/2$ satisfying (16) and $R_2=\frac{|\mathcal{I}_2|}{N}>I(X_2;Y'|X_1)$,

$$\mathbb{V}(P_{U_2^{1:N},U_1^{1:N},Y'^{1:N}},Q_{U_2^{1:N},U_1^{1:N},Y'^{1:N}})=O(2^{-N^{\beta'}}). \tag{19}$$

Note that Theorem 3 is based on the assumption that $\mathbb{V}(P_{U_1^{1:N},Y'^{1:N}},Q_{U_1^{1:N},Y'^{1:N}})=O(2^{-N^{\beta'}})$, which means that we also need $R_1>I(X_1;Y')$. Therefore, we have $\sum_{i=1}^2 R_i>I(X_1X_2;Y')$.

By induction, for level ℓ ($\ell\le r$), we define the three sets $\mathcal{F}_\ell$, $\mathcal{I}_\ell$ and $\mathcal{S}_\ell$ in the same form as (16) with $X_{1:\ell-1}^{1:N}$ replacing $X_1^{1:N}$ and U_ℓ replacing U_2. Similarly, the encoder determines $u_\ell^{1:N}(u_\ell^{1:N}=x_\ell^{1:N}G_N)$ according to the rule given by (17) and (18), with $X_{1:\ell-1}^{1:N}$ and $x_{1:\ell-1}^{1:N}$ replacing $X_1^{1:N}$ and $x_1^{1:N}$, respectively. Let $Q_{U_{1:\ell}^{1:N},Y'^{1:N}}(u_{1:\ell}^{1:N},y'^{1:N})$ denote the associate joint distribution resulted from this encoder and $P_{U_{1:\ell}^{1:N},Y'^{1:N}}(u_{1:\ell}^{1:N},y'^{1:N})$ denote the one that resulted from an encoder only using (17) for all $i\in[N]$. We have $\mathbb{V}(P_{U_{1:\ell}^{1:N},Y'^{1:N}},Q_{U_{1:\ell}^{1:N},Y'^{1:N}})=O(\ell\cdot 2^{-N^{\beta'}})$ for any rate $R_\ell=\frac{|\mathcal{I}_\ell|}{N}>I(X_\ell;Y'|X_{1:\ell-1})$. Specifically, at level r, for any rate $R_r>I(X_r;Y'|X_{1:r-1})$ and $\sum_{i=1}^r R_i>I(X_{1:r};Y')$, we have

$$\mathbb{V}(P_{U_{1:r}^{1:N},Y'^{1:N}},Q_{U_{1:r}^{1:N},Y'^{1:N}})=O(r\cdot 2^{-N^{\beta'}}). \tag{20}$$

By [11, Lemma 5], $I(X_{1:r};Y')$ is arbitrarily close to $I(X;Y')$ when N is sufficiently large and $r=O(\log\log N)$, which gives us $\mathbb{V}(P_{U_{1:r}^{1:N},Y'^{1:N}},Q_{U_{1:r}^{1:N},Y'^{1:N}})=O(2^{-N^{\beta'}})$.

Now we present the main theorem of this section. The proof is given in Appendix C.

Theorem 4: Given a Gaussian source Y with variance σ_s^2 and an average distortion $\Delta\le\sigma_s^2$, for any rate $R>\frac{1}{2}\log(\frac{\sigma_s^2}{\Delta})$, there exists a multilevel polar code with rate R such that the distortion is arbitrarily close to Δ when $N\to\infty$ and $r=O(\log\log N)$. This multilevel polar code is actually a shifted polar lattice $L+\mathrm{c}$ constructed from the lattice partition Λ/Λ' with a shaping according to the discrete Gaussian distribution D_{Λ,σ_r}, where $\sigma_r=\sqrt{\sigma_s^2-\Delta}$ and the partition chain is scaled to make $\epsilon_\Lambda(\frac{\sigma_r\sqrt{D}}{\sigma_s})\to 0$.

Remark 1. From the proof of Theorem 4, it seems that R could be slightly smaller than $\frac{1}{2}\log\frac{\sigma_s^2}{\Delta}$ (since $R>I(X;Y')\ge\frac{1}{2}\log\frac{\sigma_s^2}{\Delta}-\frac{5\epsilon_\Lambda(\tilde{\sigma}_\Delta)}{n}$) to reach an average distortion Δ, which would contradict Shannon's rate-distortion theory. However, this is not the case. When $R<\frac{1}{2}\log\frac{\sigma_s^2}{\Delta}$, an arbitrarily small $\epsilon_\Lambda(\tilde{\sigma}_\Delta)$ cannot be guaranteed, which means that the resulted distortion cannot be arbitrarily close to Δ. To achieve the desired distortion, we need $R>\frac{1}{2}\log\frac{\sigma_s^2}{\Delta}-\frac{5\epsilon_\Lambda(\tilde{\sigma}_\Delta)}{n}$ for all possibly small $\epsilon_\Lambda(\tilde{\sigma}_\Delta)$, which leads to $R>\frac{1}{2}\log\frac{\sigma_s^2}{\Delta}$.

IV. Simulation Results

The quantization performance of polar lattices for a Gaussian source with standard deviation $\sigma_s=3$ and target distortion from 0.1 to 2.5 is shown in Fig. 2. It reveals that the rate-distortion bound is approached as the dimension

of polar lattices increases from $N = 2^{10}$ to $N = 2^{18}$. Particularly, when $N = 2^{18}$, the gap to the rate-distortion bound is less than 0.2 dB.

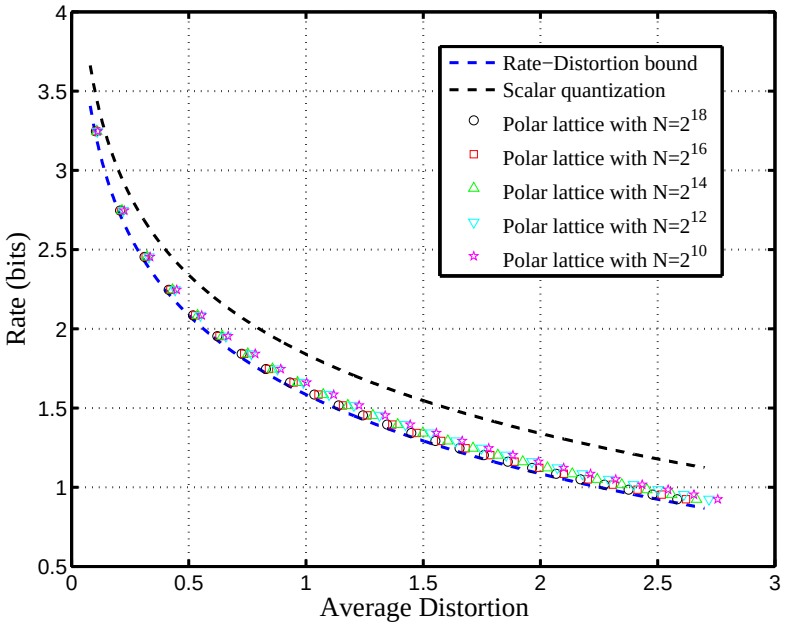
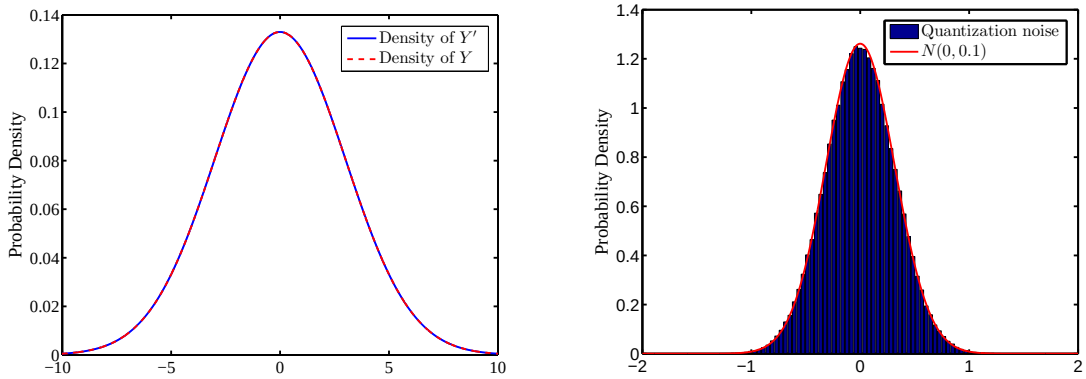


Fig. 2. Quantization performance of polar lattices for the Gaussian source with $\sigma_s = 3$.



(a) $\mathbb{V}(f_{Y'}, f_Y)$ when $r = 6$ and $\Delta = 0.5$. (b) Quantization noise v.s. Gaussian noise when $\Delta = 0.1$.

Fig. 3. Variational distance $\mathbb{V}(f_{Y'}, f_Y)$ and quantization noise for polar lattice quantizer.

In this work, the number of levels is chosen to be 6 to guarantee a negligible variational distance $\mathbb{V}(f_{Y'}, f_Y)$ for all target distortions. For a target distortion $\Delta = 0.5$, the two densities of Y' and Y are compared in Fig. 3(a), where negligible difference between f_Y and $f_{Y'}$ is found since $\mathbb{V}(f_{Y'}, f_Y) \approx 1.1 \times 10^{-7}$. Moreover, the

quantization noise behaves similarly to a Gaussian noise as shown in Fig. 3(b), which will be useful to understand the idea of Gaussian Wyner-Ziv coding and Gelfand-Pinsker coding in the next section.

A performance comparison between the TCQ and polar lattice for quantization is shown in Table I. The $\mathsf{SNR} = \frac{\sigma_s^2}{\Delta}$ is shown in dB. For the TCQ, the dimension is 1000 and the number of states is 256. For the quantization polar lattice, the dimension is 1024. It can be observed that the performance of the polar lattice is superior to that of the TCQ with roughly same block length (especially for higher rate). The performance of the Lloyd-Max scaler quantizer is also shown.

TABLE I

Performance comparison with TCQ for Gaussian source (SNR in dB).

Rate (bits)	TCQ	Polar Lattice Quantizer	Lloyd-Max Quantizer	Rate-Distortion Bound
1	5.56	5.59	4.40	6.02
2	11.04	11.55	9.30	12.04
3	16.64	17.57	14.62	18.06

V. Gaussian Wyner-Ziv Coding

A. System model

For the Wyner-Ziv problem, let X, Y be two joint Gaussian source and $X = Y + Z$, where Z is a Gaussian noise independent of Y with variance σ_z^2.[1] A typical system model of Wyner-Ziv coding for the Gaussian case is shown in Fig. 4. Given the side information Y, which is only available at the decoder's side, the Wyner-Ziv rate-distortion bound on source X for a target average distortion Δ between X and its reconstruction $\hat{X}$ is given by

$$R_{WZ}(\Delta) = \max\left\{\frac{1}{2}\log\left(\frac{\sigma_z^2}{\Delta}\right), 0\right\}. \tag{21}$$

B. A solution using continuous auxiliary variable

To achieve this bound, we assume a continuous auxiliary Gaussian random variable X' which has an average distortion Δ' with source X, i.e., $X' = X + N(0, \Delta')$. Then we can also obtain that $X' = Y + N(0, \Delta' + \sigma_z^2)$. Letting $\sigma_{x'}^2$ be the variance of X', the difference between the mutual information $I(X'; X)$ and $I(X'; Y)$ is given by

$$\begin{aligned} I(X'; X) - I(X'; Y) &= \frac{1}{2}\log\frac{\sigma_{x'}^2}{\Delta'} - \frac{1}{2}\log\frac{\sigma_{x'}^2}{\Delta' + \sigma_z^2} \\ &= \frac{1}{2}\log\frac{\Delta' + \sigma_z^2}{\Delta'}. \end{aligned} \tag{22}$$

[1]For a more general Wyner-Ziv model in the Gaussian case, the relationship between the two joint source can also be $Y = X + Z$, where $Z \sim N(0, \sigma_z^2)$ is a Gaussian noise independent of X. In this case, we can perform the MMSE rescaling on Y to make $X = \dot{\alpha}Y + \dot{Z}$, where $\dot{\alpha} = \frac{\sigma_x^2}{\sigma_y^2}$ and $\dot{Z}$ is with variance $\frac{\sigma_z^2\sigma_x^2}{\sigma_z^2+\sigma_x^2}$. Then the Wyner-Ziv bound is given by $R_{WZ}(\Delta) = \max\left\{\frac{1}{2}\log\left(\frac{\sigma_z^2\sigma_x^2}{(\sigma_z^2+\sigma_x^2)\Delta}\right), 0\right\}$. Therefore, the system model can still be described by Fig. 4, with Y and Z being replaced by $\dot{\alpha}Y$ and $\dot{Z}$, respectively.

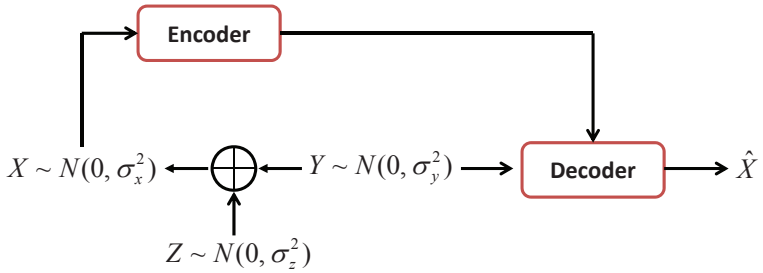


Fig. 4. Wyner-Ziv coding for the Gaussian case. The variances of Z, Y and X are given by σ_z^2, σ_y^2 and $\sigma_x^2 = \sigma_y^2 + \sigma_z^2$, respectively.

Let $I(X';X) - I(X';Y) = R_{WZ}(\Delta)$ and assume $\Delta \leq \sigma_z^2$. Then we have

$$\Delta' = \eta\Delta, \tag{23}$$

where $\eta = \frac{\sigma_z^2}{\sigma_z^2 - \Delta}$. Note that η is the reciprocal of the MMSE rescaling parameter in the scenario of quantizing a Gaussian source with variance σ_z^2 for a target average distortion Δ.

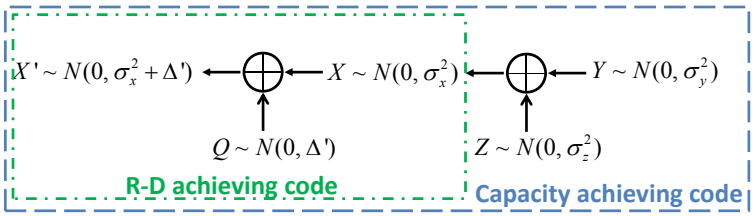


Fig. 5. A solution of the Gaussian Wyner-Ziv problem using a continuous Gaussian random variable X'.

The above-mentioned solution for the Gaussian Wyner-Ziv problem is depicted by Fig. 5. Firstly we design a lossy compression code for source X with Gaussian reconstruction alphabet X'. The average distortion between X' and X is $\Delta' = \eta\Delta$. Then we construct an AWGN capacity achieving code from Y and X'. The final reconstruction of X is given by $\hat{X} = Y + \frac{1}{\eta}(X' - Y)$. Clearly $\frac{1}{\eta}(X' - Y)$ is a scaled version of the Gaussian noise, which is independent of X'. The variance of $\frac{1}{\eta}(X' - Y)$ is

$$\begin{aligned} \frac{1}{\eta^2}(\Delta' + \sigma_z^2) &= \frac{\Delta}{\eta} + \frac{\sigma_z^2}{\eta^2} \\ &= \frac{\sigma_z^2 - \Delta}{\sigma_z^2}\Delta + \frac{\sigma_z^2 - \Delta}{\sigma_z^2}(\sigma_z^2 - \Delta) \\ &= \sigma_z^2 - \Delta. \end{aligned} \tag{24}$$

Then we can check that $X = \hat{X} + N(0, \Delta)$, which corresponds to the desired distortion, and the required data rate is $I(X';X) - I(X';Y) = \frac{1}{2}\log(\frac{\sigma_z^2}{\Delta})$.

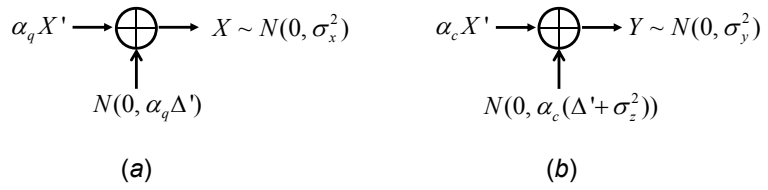


Fig. 6. The MMSE rescaled channel blocks (a) and (b) for the Gaussian channels $X \to X'$ and $Y \to X'$, respectively.

C. A practical solution using lattice Gaussian distribution

The problem of the above-mentioned solution is that X' is a continuous Gaussian random variable, which is impractical for the design of lattice codes. In order to utilize the proposed polar lattice coding technique, X' is expected to obey a lattice Gaussian distribution. To this end, we perform MMSE rescaling on X' for the AWGN channels $X \to X'$ and $Y \to X'$, respectively. The rescaled channels are shown in Fig. 6, where

$$\alpha_q = \frac{\sigma_x^2}{\sigma_x^2 + \Delta'} = \frac{\sigma_x^2(\sigma_z^2 - \Delta)}{\sigma_x^2(\sigma_z^2 - \Delta) + \sigma_z^2 \Delta}, \tag{25}$$

and

$$\alpha_c = \frac{\sigma_y^2}{\sigma_x^2 + \Delta'} = \frac{(\sigma_x^2 - \sigma_z^2)(\sigma_z^2 - \Delta)}{\sigma_x^2(\sigma_z^2 - \Delta) + \sigma_z^2 \Delta}. \tag{26}$$

Clearly, $\alpha_c < \alpha_q$. To combine the two blocks in Fig. 6 together, block (b) is scaled by $\frac{\alpha_q}{\alpha_c}$. Consequently, a reversed version of the solution illustrated in Fig. 5 is obtained and shown in Fig. 7. For the reconstruction of X, we have the following proposition.

Proposition 1: To achieve the $R_{WZ}(\Delta)$ bound by the reversed structure shown in Fig. 7, the reconstruction of X is given by

$$\hat{X} = \alpha_q X' + \gamma(\frac{\alpha_q}{\alpha_c} Y - \alpha_q X'), \quad \gamma = \frac{\sigma_y^2 \Delta}{\sigma_x^2 \sigma_z^2}. \tag{27}$$

Proof: It suffices to prove that $X = \hat{X} + N(0, \Delta)$. According to Fig. 7, we have $X = \alpha_q X' + N(0, \alpha_q \Delta')$, meaning that showing $\hat{X} = \alpha_q X' + N(0, \alpha_q \Delta' - \Delta)$ would complete this proof.

Clearly, $\frac{\alpha_q}{\alpha_c} Y - \alpha_q X'$ is a Gaussian random variable with 0 mean and variance $\alpha_q \Delta' + \frac{\alpha_q}{\alpha_c} \sigma_z^2$, and it is independent of X'. By substituting the parameters Δ', α_q and α_c, we have

$$\alpha_q \Delta' - \Delta = \frac{\sigma_x^2(\sigma_z^2 - \Delta)}{\sigma_x^2(\sigma_z^2 - \Delta) + \sigma_z^2 \Delta} \frac{\sigma_z^2}{\sigma_z^2 - \Delta} \Delta - \Delta \tag{28}$$

$$= \frac{\sigma_y^2 \Delta^2}{\sigma_x^2 \sigma_z^2 - \sigma_y^2 \Delta}, \tag{29}$$

and

$$\gamma^2(\alpha_q\Delta' + \frac{\alpha_q}{\alpha_c}\sigma_z^2) = \Big(\frac{\sigma_y^2\Delta}{\sigma_x^2\sigma_z^2}\Big)^2\Big(\frac{\sigma_x^2(\sigma_z^2-\Delta)}{\sigma_x^2(\sigma_z^2-\Delta)+\sigma_z^2\Delta}\frac{\sigma_z^2}{\sigma_z^2-\Delta}\Delta + \frac{\sigma_x^2}{\sigma_y^2}\sigma_z^2\Big), \quad (30)$$

$$= \Big(\frac{\sigma_y^2\Delta}{\sigma_x^2\sigma_z^2}\Big)^2\Big[\sigma_x^2\sigma_z^2\Big(\frac{\Delta}{\sigma_x^2\sigma_z^2-\sigma_y^2\Delta}+\frac{1}{\sigma_y^2}\Big)\Big] \quad (31)$$

$$= \frac{\sigma_y^2\Delta^2}{\sigma_x^2\sigma_z^2-\sigma_y^2\Delta}, \quad (32)$$

as we desired. □

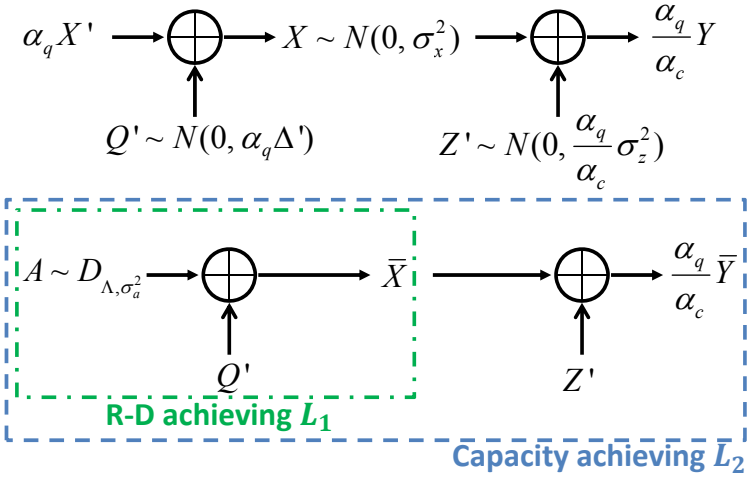


Fig. 7. A reverse solution of the Gaussian Wyner-Ziv problem, which is more compatible with lattice Gaussian distribution.

Now the continuous Gaussian random variable $\alpha_q X'$ can be replaced by a lattice Gaussian distributed variable $A \sim D_{\Lambda,\sigma_a^2}$, where $\sigma_a^2 = \alpha_q^2\sigma_{x'}^2$. Let $\bar{X} = A + N(0,\alpha_q\Delta')$ and $\frac{\alpha_q}{\alpha_c}\bar{Y} = \bar{X} + N(0,\frac{\alpha_q}{\alpha_c}\sigma_z^2)$. Let $\bar{B} = \frac{\alpha_q}{\alpha_c}\bar{Y}$ and $\sigma_b^2 = \frac{\alpha_q^2}{\alpha_c^2}\sigma_y^2$ for convenience. By Lemma 1, the distributions of $\bar{X}$ and $\bar{Y}$ can be made arbitrarily close to those of X and Y, respectively. Then the polar lattices are designed by treating $\bar{X}$ as the source and $\bar{Y}$ as its side information. A rate-distortion bound achieving polar lattice L_1 is constructed for source $\bar{X}$ with target distortion $\alpha_q\Delta'$, and an AWGN capacity-achieving polar lattice L_2 is constructed to help the decoder extract some information from $\bar{Y}$, as shown in Fig. 7. Finally, the decoder reconstructs $\check{X} = A + \gamma(\bar{B} - A)$. Conceptually, $\bar{B} - A$ is a Gaussian noise which is independent of A. [2] Recall that $\gamma = \frac{\sigma_y^2\Delta}{\sigma_x^2\sigma_z^2}$ scales $\bar{B} - A$ to $N(0,\alpha_q\Delta' - \Delta)$. By Lemma 1 again, the distributions of $\check{X}$ and $\hat{X}$ can be very close, resulting in an average distortion close to Δ.

When lattice Gaussian distribution is utilized, by [11, Lemma 10], L_1 and L_2 are accordingly constructed for

[2]In fact, when A is reconstructed by the decoder, $\bar{B} - A$ is not exactly a Gaussian noise $N(0,\alpha_q\Delta' + \frac{\alpha_q}{\alpha_c}\sigma_z^2)$, since the quantization noise of L_1 is not exactly Gaussian distributed. However, according to Theorem 4, the two distributions can be made arbitrarily close when N is sufficiently large. See Fig. 3(b) for an example.

the MMSE-rescaled Gaussian noise variance $\tilde{\sigma}_q^2$ and $\tilde{\sigma}_c^2$, where

$$\tilde{\sigma}_q^2 = \frac{\sigma_a^2}{\sigma_x^2}\alpha_q\Delta' = \frac{\sigma_a^2\sigma_z^2\Delta}{\sigma_x^2\sigma_z^2 - \sigma_y^2\Delta}, \tag{33}$$

and

$$\tilde{\sigma}_c^2 = \frac{\sigma_a^2}{\sigma_b^2}(\alpha_q\Delta' + \frac{\alpha_q}{\alpha_c}\sigma_z^2) = \frac{\sigma_a^2\sigma_z^4}{\sigma_x^2\sigma_z^2 - \sigma_y^2\Delta}. \tag{34}$$

Since $\Delta \le \sigma_z^2$, we also have $\tilde{\sigma}_q^2 \le \tilde{\sigma}_c^2$.

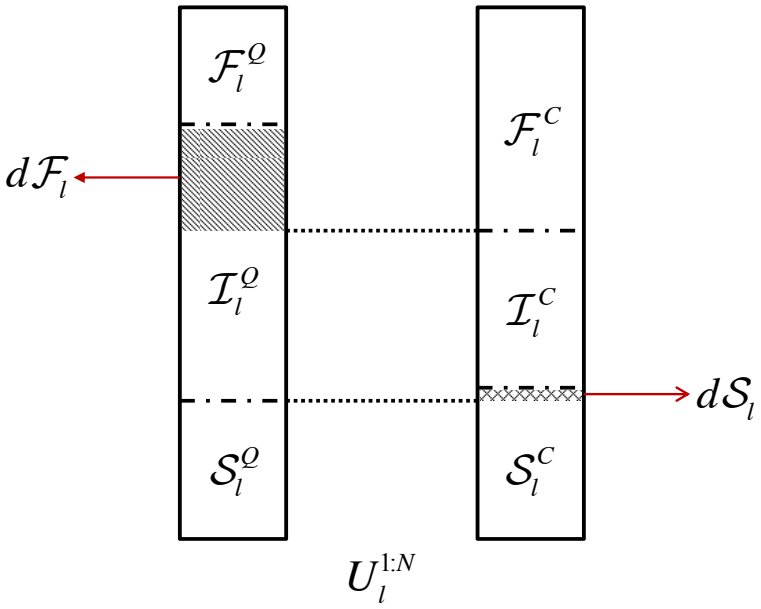


Fig. 8. The partitions of $U_\ell^{1:N}$ for quantization lattice L_1 (left) and channel coding lattice L_2 (right). $\mathcal{F}_\ell^Q \subseteq \mathcal{F}_\ell^C$, $\mathcal{I}_\ell^C \subseteq \mathcal{I}_\ell^Q$, and $\mathcal{S}_\ell^Q \subseteq \mathcal{S}_\ell^C$. Without the side information, $U_\ell^{\mathcal{I}_\ell^Q}$ should be sent to achieve the target distortion. With the side information, however, $U_\ell^{\mathcal{I}_\ell^C}$ can be decoded and hence only $U_\ell^{d\mathcal{I}_\ell}$ need to be sent.

Now we are ready to introduce the polar lattice coding strategy. The idea is similar to the one mentioned in [15]. We choose a good constellation D_{Λ,σ_a^2} such that the flatness factor $\epsilon_\Lambda(\tilde{\sigma}_q)$ is negligible. Let $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'/\cdots$ be a one-dimensional binary partition chain labeled by bits $A_1/A_2/\cdots/A_{r-1}/A_r/\cdots$. Then $P_{A_{1:r}}$ and $A_{1:r}$ approach D_{Λ,σ_a^2} and A, respectively, as $r \to \infty$. Consider N i.i.d. copies of A. Let $U_\ell^{1:N} = A_\ell^{1:N}G_N$ for each $1 \le \ell \le r$. The partitions of $U_\ell^{1:N}$ for both L_1 and L_2 are shown in Fig. 8, where the left block is for quantization lattice L_1 and the right one for channel coding lattice L_2. According to Section III and [11], for $0 < \beta < 0.5$, the frozen set $\mathcal{F}_\ell^Q$ ($\mathcal{F}_\ell^C$), information set $\mathcal{I}_\ell^Q$ ($\mathcal{I}_\ell^C$) and the shaping set $\mathcal{S}_\ell^Q$ ($\mathcal{S}_\ell^C$) for lattice L_1 (L_2) are given by

$$\begin{cases} \mathcal{F}_\ell^Q = \{i \in [N] : Z(U_\ell^i|U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{X}^{1:N}) \ge 1 - 2^{-N^\beta}\} \\ \mathcal{I}_\ell^Q = \{i \in [N] : Z(U_\ell^i|U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) > 2^{-N^\beta} \text{ and } Z(U_\ell^i|U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{X}^{1:N}) < 1 - 2^{-N^\beta}\} \\ \mathcal{S}_\ell^Q = \{i \in [N] : Z(U_\ell^i|U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) \le 2^{-N^\beta}\}, \end{cases} \tag{35}$$

and

$$\begin{cases} \mathcal{F}_\ell^C = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{B}^{1:N}) \geq 1 - 2^{-N^\beta}\} \\ \mathcal{I}_\ell^C = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) \geq 1 - 2^{-N^\beta} \text{ and } Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{B}^{1:N}) \leq 2^{-N^\beta}\} \\ \mathcal{S}_\ell^C = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) < 1 - 2^{-N^\beta} \text{ or } 2^{-N^\beta} < Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{B}^{1:N}) < 1 - 2^{-N^\beta}\}. \end{cases} \tag{36}$$

For these two partitions, we have the following lemma.

Lemma 2: Let L_1 and L_2 be two polar lattices constructed according to the above two partition rules respectively. L_2 is nested within L_1, i.e., $L_2 \subseteq L_1$.

Proof: Both L_1 and L_2 follow the multilevel lattice structure (4). Let $\{C_1^q, ..., C_\ell^q, ..., C_r^q\}$ and $\{C_1^c, ..., C_\ell^c, ..., C_r^c\}$ denote the multilevel codes for L_1 and L_2, respectively. When shaping is not involved, the generator matrixes of C_ℓ^q and C_ℓ^c correspond to the sets of row indices $\mathcal{I}_\ell^Q \cup \mathcal{S}_\ell^Q$ and $\mathcal{I}_\ell^C \cup \mathcal{S}_\ell^Q$, respectively. By the relationship $\tilde{\sigma}_q^2 \leq \tilde{\sigma}_c^2$ and [11, Lemma 3], the partition channel $C(\Lambda_{\ell-1}/\Lambda_\ell, \tilde{\sigma}_c^2)$ is degraded with respect to $C(\Lambda_{\ell-1}/\Lambda_\ell, \tilde{\sigma}_q^2)$. Then by the equivalence lemma [11, Lemma 10], we have $\mathcal{I}_\ell^C \subseteq \mathcal{I}_\ell^Q$, meaning that $C_\ell^c \subseteq C_\ell^q$ for $1 \leq \ell \leq r$. As a result, $L_2 \subseteq L_1$. □

By channel degradation, we have $\mathcal{F}_\ell^Q \subseteq \mathcal{F}_\ell^C$. Let $d\mathcal{F}_\ell$ denote the set $\mathcal{F}_\ell^C \setminus \mathcal{F}_\ell^Q$. Meanwhile, we have $\mathcal{S}_\ell^Q \subseteq \mathcal{S}_\ell^C$ by definition. Denoting by $d\mathcal{S}_\ell$ the set $\mathcal{S}_\ell^C \setminus \mathcal{S}_\ell^Q$, $d\mathcal{S}_\ell$ can be written as

$$\begin{aligned} d\mathcal{S}_\ell = \{i \in [N] : 2^{-N^\beta} < Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) < 1 - 2^{-N^\beta} \text{ or} \\ 2^{-N^\beta} < Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{B}^{1:N}) < 1 - 2^{-N^\beta}\}, \end{aligned} \tag{37}$$

and the proportion $\frac{|d\mathcal{S}_\ell|}{N} \to 0$ as $N \to \infty$. Also observe that $d\mathcal{I}_\ell = \mathcal{I}_\ell^Q \setminus \mathcal{I}_\ell^C = d\mathcal{F}_\ell \cup d\mathcal{S}_\ell$.

Given an N-dimensional realization vector $x^{1:N}$ of $X^{1:N}$, the encoder evaluates $u_\ell^{1:N}$ from level 1 to level r successively according to the random rounding quantization rules given in Section III. (See (13), (14), (17) and (18).) Recall that treating $x^{1:N}$ as a realization of $\bar{X}^{1:N}$ is safe because X and $\bar{X}$ are similarly distributed. Then $u_\ell^{d\mathcal{I}_\ell}$ is sent to the decoder for each level. For the decoder, the realization vector $y^{1:N}$ of $Y^{1:N}$ is scaled to $b^{1:N} = \frac{\alpha_q}{\alpha_c} y^{1:N}$. Since $u_\ell^{\mathcal{F}_\ell^Q}$ is shared between the encoder and decoder before transmission, after receiving $u_\ell^{d\mathcal{I}_\ell}$, $u_\ell^{\mathcal{I}_\ell^C}$ and $u_\ell^{\mathcal{S}_\ell^Q}$ can be decoded with vanishing error probability since their associate Bhattacharyya parameters are arbitrarily small when $N \to \infty$. The details of SC decoding for Gaussian channels have been discussed in [11]. According to [11, Lemma 8], probabilities $P_{U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}}$, $P_{U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{X}^{1:N}}$ and $P_{U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{B}^{1:N}}$ can be evaluated with $O(N \log N)$ complexity. It is worth mentioning that $u_\ell^{\mathcal{S}_\ell^C}$ is covered by a pre-shared random mapping in [11]. However, as shown in Theorem 2, replacing the random mapping with MAP decision for $u_\ell^{\mathcal{S}_\ell^Q}$ will not change the results of [11, Theorem 5] and [11, Theorem 6]. The proof is similar to that of Theorem 2 and omitted for brevity. Then the whole vector $u_\ell^{1:N}$ can be recovered with high probability. After obtaining the $u_\ell^{1:N}$ for $1 \leq \ell \leq r$, the realization $a^{1:N}$ of $A^{1:N}$ can be recovered from $u_\ell^{1:N}$ according to the following equation

$$\chi = \sum_{\ell=1}^{r} 2^{\ell-1} \left[\sum_{i \in \mathcal{I}_\ell} u_\ell^i \psi(\mathbf{g}_i) + \sum_{i \in \mathcal{S}_\ell} u_\ell^i \psi(\mathbf{g}_i) + \sum_{i \in \mathcal{F}_\ell} u_\ell^i \psi(\mathbf{g}_i) \right], \tag{38}$$

where $\mathbf{g}_i$ denotes the i-th row of the polarization matrix G_N and ψ is the natural embedding. Clear $a^{1:r}$ is drawn from $D_{2^r\mathbb{Z}^N+\chi,\sigma_a}$. For each dimension, when r is sufficiently large, the probability of choosing a constellation point outside the interval $[-2^{r-1}, 2^{r-1}]$ is negligible (see [11, Lemma 5] for more detail). Therefore, there exists only one point within $[-2^{r-1}, 2^{r-1}]$ with probability close to 1 and $a^{1:N}$ can be recovered by $\chi \mod 2^r$. Finally, the reconstruction of $x^{1:N}$ is given by $\breve{x}^{1:N} = a^{1:N} + \gamma(b^{1:N} - a^{1:N})$.

To sum up, we have the following Wyner-Ziv coding scheme.

- Encoding: For the N-dimensional i.i.d. source vector $X^{1:N}$, the encoder evaluates $U_\ell^{\mathcal{I}_\ell^Q}$ by random rounding, and then sends $U_\ell^{d\mathcal{I}_\ell}$ to the decoder.
- Decoding: Using the pre-shared $U^{\mathcal{F}_\ell^Q}$ and the received $U_\ell^{d\mathcal{I}_\ell}$, the decoder recovers $U^{\mathcal{I}_\ell^C}$ and $U^{\mathcal{S}_\ell^Q}$ from the side information $B^{1:N}$. For each level the decoder obtains $U_\ell^{1:N}$, then $A^{1:N}$ can be recovered according to (38).
- Reconstruction: $\tilde{X}^{1:N} = A^{1:N} + \gamma(B^{1:N} - A^{1:N})$.

With regard to the design rate, by Theorem 4, the rate R_{L_1} of L_1 can be arbitrarily close to $\frac{1}{2}\log\frac{\sigma_x^2}{\alpha_q\Delta'}$. However, the encoder does not need to send that much information to the decoder because of the side information. By [11, Theorem 7], the rate R_{L_2} of L_2 can be arbitrarily close to $\frac{1}{2}\log\big(\frac{\sigma_b^2}{\alpha_q\Delta'+\frac{\alpha_q}{\alpha_c}\sigma_z^2}\big)$. After some tedious calculation, we have

$$R_{L_1} \to \frac{1}{2}\log\left(\frac{\sigma_x^2\sigma_z^2 - \sigma_y^2\Delta}{\sigma_z^2\Delta}\right)^+, \tag{39}$$

and

$$R_{L_2} \to \frac{1}{2}\log\left(\frac{\sigma_x^2\sigma_z^2 - \sigma_y^2\Delta}{\sigma_z^4}\right)^-, \tag{40}$$

meaning that the transmission rate $R_{L_1} - R_{L_2} \to \frac{1}{2}\log(\frac{\sigma_z^2}{\Delta})^+$.

Before presenting the main theorem of the Gaussian Wyner-Ziv coding, we need a more stringent requirement on the flatness factor. This requirement is to guarantee a sub-exponentially decaying error probability for our lattice coding scheme.

Proposition 2: For a one-dimensional binary partition chain $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'/\cdots$ and any given $\tilde{\sigma}$, $r = O(\log N)$ is sufficient to guarantee a sub-exponentially vanishing flatness factor $\epsilon_\Lambda(\tilde{\sigma}) = O(2^{-\sqrt{N}})$. Moreover, the mutual information of the bottom level $I(\bar{X}; A_r|A_{1:r-1}) \to 0$ and using the first r levels only incurs a capacity loss $\sum_{\ell>r} I(\bar{X}; A_\ell|A_{1:\ell-1}) \le O(\frac{1}{N})$.

Proof: Since the partition is with dimension one, we can assume that $\Lambda = \eta\mathbb{Z}$. Let $\Lambda^* = \frac{1}{\eta}\mathbb{Z}$ be the dual lattice of Λ. By [24, Corollary 1], we have

$$\begin{aligned}
\epsilon_\Lambda(\tilde{\sigma}) &= \Theta_{\Lambda^*}(2\pi\tilde{\sigma}^2) - 1 && (41)\\
&= \sum_{\lambda\in\Lambda^*} \exp(-2\pi^2\tilde{\sigma}^2\|\lambda\|^2) - 1 && (42)\\
&= 2\sum_{\lambda\in\frac{1}{\eta}\mathbb{Z}_+} \exp(-2\pi^2\tilde{\sigma}^2\|\lambda\|^2) && (43)
\end{aligned}$$

$$\leq \frac{2\exp(-2\pi^2\tilde{\sigma}^2\frac{1}{\eta^2})}{1-\exp(-2\pi^2\tilde{\sigma}^2\frac{3}{\eta^2})} \tag{44}$$

$$\leq 4\exp(-2\pi^2\tilde{\sigma}^2\frac{1}{\eta^2}), \tag{45}$$

where Z_+ denotes positive integers and the last inequality satisfies for sufficiently small η. Let $\frac{1}{\eta^2} = O(\sqrt{N})$ and hence $\epsilon_\Lambda(\tilde{\sigma}) = O(2^{-\sqrt{N}})$. In addition, by [11, Lemma 2], a number of levels $r_1 = O(\log\log N)$ is needed to guarantee a vanishing mutual information at the bottom level. Let $\delta\mathbb{Z}/\cdots/2^{r_1}\mathbb{Z}$ be a partition such that $I(\bar{X}; A_r|A_{1:r-1}) \to 0$ for a constant δ. Finally, the number of levels for partition $\eta\mathbb{Z}/\cdots/\delta\mathbb{Z}/\cdots/2^{r_1}\mathbb{Z}$ satisfies $r = \log(\frac{2^{r_1}}{\eta}) = O(\log N)$. □

The following theorem is proved in Appendix D.

Theorem 5: Let X be a Gaussian source and Y be another Gaussian source correlated to X as $X = Y + Z$, where $Z \sim N(0, \sigma_z^2)$ is an independent Gaussian noise. Consider a target distortion $0 \leq \Delta \leq \sigma_z^2$ for source X when Y is only available for the decoder. Let $\Lambda/\Lambda_1/\cdots/\Lambda_r$ be a one-dimensional binary partition chain such that $\epsilon_\Lambda(\tilde{\sigma}_q) = O(2^{-\sqrt{N}})$ and $r = O(\log N)$. For any $0 < \beta' < \beta < 0.5$, there exists two nested polar lattices L_1 and L_2 with a differential rate $R = R_{L_1} - R_{L_2}$ arbitrarily close to $\frac{1}{2}\log(\frac{\sigma_z^2}{\Delta})$ such that the expect distortion Δ_Q satisfies

$$\Delta_Q \leq \Delta + O(2^{-N^{\beta'}}), \tag{46}$$

and the block error probability satisfies

$$P_e^{WZ} \leq O(2^{-N^{\beta'}}). \tag{47}$$

VI. Gaussian Gelfand-Pinsker coding

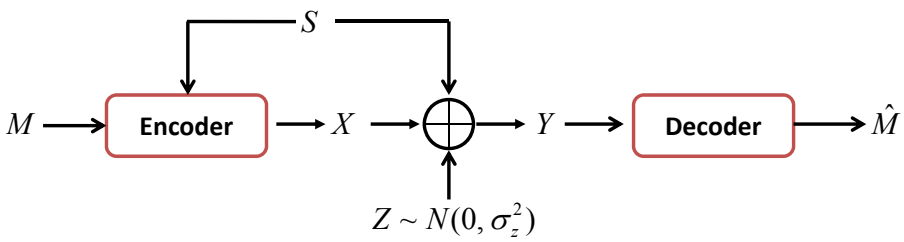


Fig. 9. Gelfand-Pinsker coding for the Gaussian case.

For the Gelfand-Pinsker problem, with some abuse of notations, consider the channel described by $Y = X+S+Z$, where X and Y are the channel input and output, respectively, Z is an unknown additive Gaussian noise with variance σ_z^2 and S is an interference Gaussian signal with variance σ_i^2 known only to the encoder. A diagram of Gelfand-Pinsker coding is shown in Fig. 9. Message M is encoded into X which satisfies the power constraint $\frac{1}{N}E[\|X^{1:N}\|^2] \leq P$. The channel capacity of this Gaussian Gelfand-Pinsker model [31], [32] is given by

$$C_{GP} = \frac{1}{2}\log\left(1 + \frac{P}{\sigma_z^2}\right).$$

To achieve this capacity, the roles of quantization lattice and channel coding lattice are reversed. To see this, we still start with a continuous auxiliary variable and then replace it with a discrete Gaussian distributed one. Letting $\rho = \frac{P}{P+\sigma_z^2}$, we firstly design a lossy compression code for ρS with Gaussian reconstruction alphabet S'. The distortion between S' and ρS is targeted to be P, i.e., $S' = \rho S + N(0, P)$. Then the encoder transmits $X = S' - \rho S$ (X is independent of S), which satisfies the power constraint. Moreover, the relationship between Y and S' is given by

$$\begin{aligned} S' &= X + \rho S \\ &= X + \rho(Y - X - Z) \\ &= \rho Y + (1-\rho)X - \rho Z. \end{aligned} \tag{48}$$

Then the expectation

$$E\big[Y \cdot [(1-\rho)X - \rho Z]\big] = (1-\rho)E[X^2] - \rho E[Z^2] = 0, \tag{49}$$

meaning that $(1-\rho)X - \rho Z$ is independent of Y, which gives $S' = \rho Y + N(0, \frac{P\sigma_z^2}{P+\sigma_z^2})$. Then we construct an AWGN capacity-achieving code to recover S' from ρY. Without the power constraint, the maximum data rate that can be sent is actually $I(S'; \rho Y)$. However, when power constraint is taken into consideration, some bits should be selected according to the realization of S since S' and S are related. The maximum data rate becomes $I(S'; \rho Y) - I(S'; \rho S) = \frac{1}{2}\log(1 + \frac{P}{\sigma_z^2}) = C_{GP}$. A diagram of this solution is shown in Fig. 10, where

$$\sigma_{s'}^2 = \rho^2 \sigma_i^2 + P, \tag{50}$$

and

$$\sigma_y^2 = \frac{1}{\rho^2}\frac{P^2}{P+\sigma_z^2} + \sigma_i^2 = \sigma_i^2 + P + \sigma_z^2 \tag{51}$$

are the variances of S' and Y, respectively.

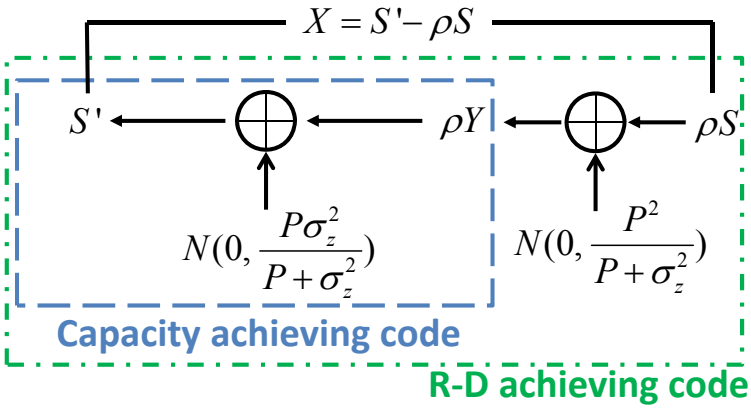


Fig. 10. A solution of the Gaussian Gelfand-Pinsker problem using continuous Gaussian random variable S'.

Similarly, we prefer to use a discrete lattice Gaussian distributed version of S' to approach this capacity. The idea is to perform MMSE rescaling on S' to get a reversed version of the model shown in Fig. 10. The analysis

is similar to that presented in Section V and is omitted here for brevity. Finally, the reversed solution is given in Fig. 11.

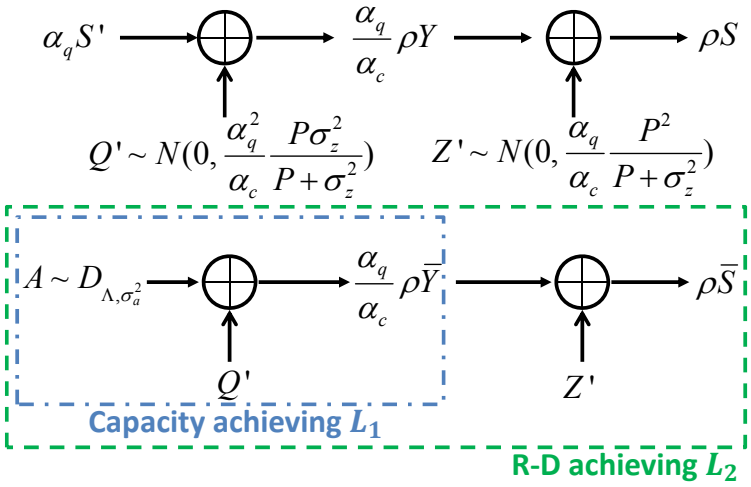


Fig. 11. A reverse solution of the Gaussian Gelfand-Pinsker problem.

With some abuse of notation, let A denote the discrete version of $\alpha_q S'$. The MMSE rescaling factor α_c for channel coding and α_q for quantization are given by

$$\alpha_c = \frac{P\sigma_y^2}{P\sigma_i^2 + (P+\sigma_z^2)^2}, \tag{52}$$

and

$$\alpha_q = \frac{P\sigma_i^2}{P\sigma_i^2 + (P+\sigma_z^2)^2}, \tag{53}$$

respectively. The variance σ_a^2 for D_{Λ,σ_a^2} is chosen to be $\alpha_q^2\sigma_{s'}^2$. Polar lattices L_1 and L_2 are accordingly constructed for Gaussian noise variance $\tilde{\sigma}_c^2$ and $\tilde{\sigma}_q^2$, where

$$\tilde{\sigma}_c^2 = \frac{\alpha_q^2\alpha_c}{\rho} \cdot \frac{\sigma_z^2\sigma_{s'}^2}{\sigma_y^2}, \tag{54}$$

and

$$\tilde{\sigma}_q^2 = \frac{\alpha_q^3}{\rho^2} \cdot \frac{P\sigma_{s'}^2}{\sigma_i^2}. \tag{55}$$

Check that $\frac{\tilde{\sigma}_c^2}{\tilde{\sigma}_q^2} = \frac{\sigma_z^2}{P+\sigma_z^2} \leq 1$. Recall that $X = S' - \rho S = (1-\alpha_q)S' + \alpha_q S' - \rho S$. When $\alpha_q S'$ is replaced by A, the encoded signal, denoted by $\bar{X}$, is given by

$$\bar{X} = \frac{1-\alpha_q}{\alpha_q}A + A - \rho S. \tag{56}$$

Note that the distributions of S and $\bar{S}$ can be arbitrarily close when $\epsilon_\Lambda(\tilde{\sigma}_c) \to 0$. Clearly, $A - \rho\bar{S}$ is a Gaussian random variable independent of A with distribution $N(0, \alpha_q P)$. By Lemma 1, $\bar{X}$ can be very close to a Gaussian random variable with distribution $N(0, P)$. [3] Thus, the power constraint can be satisfied.

[3]Check that $\left(\frac{1-\alpha_q}{\alpha_q}\right)^2\sigma_a^2 = (1-\alpha_q)P$.

With some abuse of notations, let $B = \frac{\alpha_q}{\alpha_c}\rho Y$, $\bar{B} = \frac{\alpha_q}{\alpha_c}\rho \bar{Y}$, $T = \rho S$, and $\bar{T} = \rho \bar{S}$ for convenience. We choose a good constellation D_{Λ,σ_a^2} such that the flatness factor $\epsilon_\Lambda(\tilde{\sigma}_c)$ is negligible. Let $\Lambda/\Lambda_1/\cdots/\Lambda_{r-1}/\Lambda'/\cdots$ be a one-dimensional binary partition chain labeled by bits $A_1/A_2/\cdots/A_{r-1}/A_r/\cdots$. Then $P_{A_{1:r}}$ and $A_{1:r}$ approach D_{Λ,σ_a^2} and A, respectively, as $r \to \infty$. Consider N i.i.d. copies of A. Let $U_\ell^{1:N} = A_\ell^{1:N} G_N$ for each $1 \le \ell \le r$. The partition of $U_\ell^{1:N}$ is shown in Fig. 12, where the left block is for quantization lattice L_2 and the right one for channel coding lattice L_1. For $0 < \beta < 0.5$, the frozen set $\mathcal{F}_\ell^Q$ ($\mathcal{F}_\ell^C$), information set $\mathcal{I}_\ell^Q$ ($\mathcal{I}_\ell^C$) and the shaping set $\mathcal{S}_\ell^Q$ ($\mathcal{S}_\ell^C$) for lattice L_2 (L_1) are given by

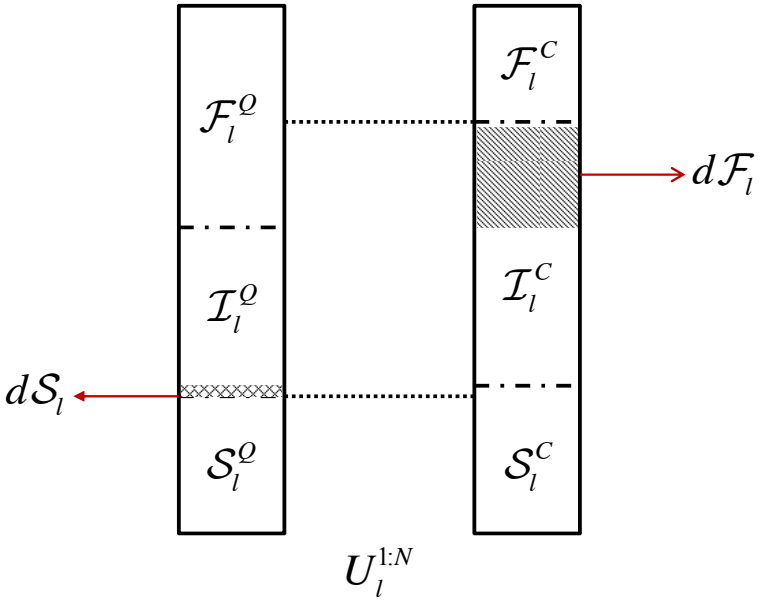


Fig. 12. The partitions of $U_\ell^{1:N}$ for quantization lattice L_2 (left) and channel coding lattice L_1 (right). $\mathcal{F}_\ell^C \subseteq \mathcal{F}_\ell^Q$ and $\mathcal{S}_\ell^Q \subseteq \mathcal{S}_\ell^C$. Without the power constraint, $U_\ell^{\mathcal{I}_\ell^C}$ can be sent as message bits. With the power constraint, however, $U_\ell^{\mathcal{I}_\ell^Q}$ should be selected according to the interference $S^{1:N}$ and hence only $U_\ell^{d\mathcal{F}_\ell}$ can be fed with the message bits.

$$\begin{cases} \mathcal{F}_\ell^Q = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{T}^{1:N}) \ge 1 - 2^{-N^\beta}\} \\ \mathcal{I}_\ell^Q = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) > 2^{-N^\beta} \text{ and } Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{T}^{1:N}) < 1 - 2^{-N^\beta}\} \\ \mathcal{S}_\ell^Q = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) \le 2^{-N^\beta}\}, \end{cases} \tag{57}$$

and

$$\begin{cases} \mathcal{F}_\ell^C = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{B}^{1:N}) \ge 1 - 2^{-N^\beta}\} \\ \mathcal{I}_\ell^C = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) \ge 1 - 2^{-N^\beta} \text{ and } Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{B}^{1:N}) \le 2^{-N^\beta}\} \\ \mathcal{S}_\ell^C = \{i \in [N] : Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}) < 1 - 2^{-N^\beta} \text{ or } 2^{-N^\beta} < Z(U_\ell^i | U_\ell^{1:i-1}, A_{1:\ell-1}^{1:N}, \bar{B}^{1:N}) < 1 - 2^{-N^\beta}\}. \end{cases} \tag{58}$$

By channel degradation, we have $\mathcal{F}_\ell^C \subseteq \mathcal{F}_\ell^Q$. Let $d\mathcal{F}_\ell$ denote the set $\mathcal{F}_\ell^Q \setminus \mathcal{F}_\ell^C$. Meanwhile, we also have $\mathcal{S}_\ell^Q \subseteq \mathcal{S}_\ell^C$. The difference $d\mathcal{S}_\ell = \mathcal{S}_\ell^C \setminus \mathcal{S}_\ell^Q$ can also be written as (37), and the proportion $\frac{|d\mathcal{S}_\ell|}{N} \to 0$ as $N \to \infty$.

Given an N-dimensional realization vector $s^{1:N}$ of $S^{1:N}$, the encoder scales $s^{1:N}$ to $t^{1:N} = \rho s^{1:N}$ and evaluates $u_\ell^{1:N}$ from level 1 to level r successively according to the random rounding quantization rules. Note that $u_\ell^{\mathcal{F}_\ell^C}$ is uniformly random and known to the decoder, and $u_\ell^{d\mathcal{F}_\ell}$ is fed with message bits which are also uniform. Recall that treating $t^{1:N}$ as a realization of $\bar{T}^{1:N}$ is reasonable because T and $\bar{T}$ are similarly distributed. Then $u_\ell^{1:N}$ can be obtained for $1 \le \ell \le r$. When r is sufficiently large, the lattice points outside $[-2^{r-1}, 2^{r-1}]$ occur with almost 0 probability, the realization $a^{1:N}$ of $A^{1:N}$ can be determined by $u_{1:r}^{1:N}$ according to (38). Then $x^{1:N} = \frac{1}{\alpha_q} a^{1:N} - \rho s^{1:N}$ is the encoded signal as discussed in (56).

For the encoder, the realization vector $y^{1:N}$ of $Y^{1:N}$ is scaled to $b^{1:N} = \frac{\alpha_q}{\alpha_c} \rho y^{1:N}$. The task is to recover $u_\ell^{\mathcal{I}_\ell^C}$ at each level and hence message $u_\ell^{d\mathcal{F}_\ell}$ can be obtained. Note that $u_\ell^{\mathcal{F}_\ell^Q}$ is shared between the encoder and decoder before transmission, and $u_\ell^{\mathcal{S}_\ell^Q}$ can be decoded with vanishing error probability using the bit-wise MAP rule. The decoder still needs to know the unpolarized bits $u_\ell^{d\mathcal{S}_\ell}$ since the Bhattacharyya parameters of those indices are not necessarily vanishing. Therefore, a code with negligible rate is needed to send $u_\ell^{\mathcal{S}_\ell^Q}$ to the decoder in advance at each level. This two phases transmission method has already been used in [15]. In this sense, L_2 is not exactly nested within L_1 because of those unpolarized indices. When $u_\ell^{d\mathcal{S}_\ell}$ is also available, $u_\ell^{\mathcal{I}_\ell^C}$ can be decoded with very small error probability [11] with $O(N \log N)$ complexity.

The Gaussian Gelfand-Pinsker coding scheme is summarized as follows.

- Encoding: According to the N-dimensional i.i.d. interference vector $S^{1:N}$, the encoder evaluates $U^{\mathcal{I}_\ell^Q}$ by random rounding, and then feeds $U_\ell^{d\mathcal{F}_\ell}$ with message bits. $U^{\mathcal{F}_\ell^C}$ is pre-shared and $U^{\mathcal{S}_\ell^Q}$ is determined by other bits according to D_{Λ,σ_a^2}. For each level, the encoder obtains $U_\ell^{1:N}$ for $1 \le \ell \le r$, then $A^{1:N}$ is recovered from $U_{1:r}^{1:N}$. The encoded signal is given by
$$\bar{X}^{1:N} = \frac{1}{\alpha_q} A^{1:N} - \rho S^{1:N}. \tag{59}$$
- Decoding: Using the pre-shared $U^{\mathcal{F}_\ell^Q}$ and the bits $U_\ell^{d\mathcal{S}_\ell}$ by the two phases transmission, the decoder recovers $U^{\mathcal{I}_\ell^C}$ including the message bits and $U^{\mathcal{S}_\ell^Q}$ from the received signal.

For the rate of lattice codes, we have
$$R_{L_1} \to \frac{1}{2} \log \left(\frac{P\sigma_i^2 + (P+\sigma_z^2)^2}{\sigma_z^2 (P+\sigma_z^2)} \right)^-, \tag{60}$$
and
$$R_{L_2} \to \frac{1}{2} \log \left(\frac{P\sigma_i^2 + (P+\sigma_z^2)^2}{(P+\sigma_z^2)^2} \right)^+, \tag{61}$$
indicating that the $R_{L_1} - R_{L_2} \to C_{GP}$.

The proof of the following theorem is given in Appendix E.

Theorem 6: Let S be a Gaussian noise known to the encoder, and Z be another independent and unknown Gaussian noise with variance σ_z^2. Consider a power constraint P for the encoded signal. Let $\Lambda/\Lambda_1/\cdots/\Lambda_r$ be a one-dimensional binary partition chain such that $\epsilon_\Lambda(\tilde{\sigma}_c) = O(2^{-\sqrt{N}})$ and $r = O(\log N)$. For any $0 < \beta' < \beta < 0.5$,

there exists two nested polar lattices L_1 and L_2 with a differential rate $R = R_{L_1} - R_{L_2}$ arbitrarily close to $\frac{1}{2}\log(1+\frac{P}{\sigma_z^2})$ such that the expect transmit power P_T satisfies

$$P_T \leq P + O(2^{-N^{\beta'}}), \tag{62}$$

and the block error probability satisfies

$$P_e^{GP} \leq O(2^{-N^{\beta'}}). \tag{63}$$

VII. Concluding Remarks

In this work, we present an explicit construction of polar lattices which are good for lossy compression. They are further utilized to resolve the Gaussian version of the Wyner-Ziv and Gelfand-Pinsker problems. Compared with the original idea given in [19], dither is not necessary in our scheme due to the property of discrete lattice Gaussian distribution [25], and the entropy encoder is already integrated in our lattice quantization process. The complexity of encoding and decoding complexity is $O(N\log^2 N)$ for a sub-exponentially decaying excess distortion.

Appendix A
Proof of Theorem 2

Proof: Firstly, we change the encoding rule for the u_1^i in $i \in \mathcal{S}_1$ and (14) is modified to

$$u_1^i = \begin{cases} \bar{u}_1^i & \text{if} \quad i \in \mathcal{F}_1 \\ \begin{cases} 0 & \text{w. p.} \quad P_{U_1^i|U_1^{1:i-1}}(0|u_1^{1:i-1}) \\ 1 & \text{w. p} \quad P_{U_1^i|U_1^{1:i-1}}(1|u_1^{1:i-1}) \end{cases} & \text{if} \quad i \in \mathcal{S}_1. \end{cases} \tag{64}$$

Let $Q'_{U_1^{1:N},Y'^{1:N}}(u_1^{1:N}, y'^{1:N})$ denote the associate joint distribution for $U_1^{1:N}$ and $Y_1^{1:N}$ according to the encoding rule described in (13) and (64). Then the variational distance between $P_{U^{1:N},Y^{1:N}}$ and $Q'_{U^{1:N},Y^{1:N}}$ can be bounded

as follows.

$$
\begin{aligned}
&2\mathbb{V}(P_{U_1^{1:N},Y'^{1:N}}, Q'_{U_1^{1:N},Y'^{1:N}}) \\
&= \sum_{u_1^{1:N},y'^{1:N}} |Q'(u_1^{1:N},y'^{1:N}) - P(u_1^{1:N},y'^{1:N})| \\
&\overset{(a)}{=} \sum_{u_1^{1:N},y'^{1:N}} |\sum_i (Q'(u_1^i|u_1^{1:i-1},y'^{1:N}) - P(u_1^i|u_1^{1:i-1},y'^{1:N}))(\prod_{j=1}^{i-1} P(u_1^i|u_1^{1:i-1},y'^{1:N}))(\prod_{j=i+1}^{N} Q(u_1^i|u_1^{1:i-1},y'^{1:N}))P(y'^{1:N})| \\
&\overset{(b)}{\leq} \sum_{i\in\mathcal{F}_1\cup\mathcal{S}_1} \sum_{u_1^{1:N},y'^{1:N}} |Q(u_1^i|u_1^{1:i-1},y'^{1:N}) - P(u_1^i|u_1^{1:i-1},y'^{1:N})|(\prod_{j=1}^{i-1} P(u_1^i|u_1^{1:i-1},y'^{1:N}))(\prod_{j=i+1}^{N} Q(u_1^i|u_1^{1:i-1},y'^{1:N}))P(y'^{1:N}) \\
&= \sum_{i\in\mathcal{F}_1\cup\mathcal{S}_1} \sum_{u_1^{1:i},y'^{1:N}} |Q(u_1^i|u_1^{1:i-1},y'^{1:N}) - P(u_1^i|u_1^{1:i-1},y'^{1:N})|(\prod_{j=1}^{i-1} P(u_1^i|u_1^{1:i-1},y'^{1:N}))P(y'^{1:N}) \\
&= \sum_{i\in\mathcal{F}_1\cup\mathcal{S}_1} \sum_{u_1^{1:i-1},y'^{1:N}} 2P(u_1^{1:i-1},y'^{1:N})\mathbb{V}(Q_{U^i|U_1^{1:i-1}=u_1^{1:i-1},Y'^{1:N}=y'^{1:N}}, P_{U_1^i|U_1^{1:i-1}=u_1^{1:i-1},Y'^{1:N}=y'^{1:N}}) \\
&\overset{(c)}{\leq} \sum_{i\in\mathcal{F}_1\cup\mathcal{S}_1} \sum_{u_1^{1:i-1},y'^{1:N}} P(u_1^{1:i-1},y'^{1:N})\sqrt{2\ln 2 D(P_{U_1^i|U_1^{1:i-1}=u_1^{1:i-1},Y'^{1:N}=y'^{1:N}}||Q_{U_1^i|U_1^{1:i-1}=u_1^{1:i-1},Y'^{1:N}=y'^{1:N}})} \\
&\overset{(d)}{\leq} \sum_{i\in\mathcal{F}_1\cup\mathcal{S}_1} \sqrt{2\ln 2 \sum_{u_1^{1;i-1},y'^{1:N}} P(u_1^{1:i-1},y'^{1:N}) D(P_{U_1^i|U_1^{1:i-1}=u_1^{1:i-1},Y'^{1:N}=y'^{1:N}}||Q_{U_1^i|U_1^{1:i-1}=u_1^{1:i-1},Y'^{1:N}=y'^{1:N}})} \\
&\leq \sum_{i\in\mathcal{F}_1\cup\mathcal{S}_1} \sqrt{2\ln 2 D(P_{U_1^i}||Q_{U_1^i}|U_1^{1:i-1},Y'^{1:N})} \\
&\overset{(e)}{\leq} \sum_{i\in\mathcal{F}_1} \sqrt{2\ln 2(1-H(U_1^i|U_1^{1:i-1},Y'^{1:N}))} + \sum_{i\in\mathcal{S}_1} \sqrt{2\ln 2(H(U_1^i|U_1^{1:i-1}) - H(U_1^i|U_1^{1:i-1},Y'^{1:N}))} \\
&\overset{(f)}{\leq} \sum_{i\in\mathcal{F}_1} \sqrt{2\ln 2(1-Z(U_1^i|U_1^{1:i-1},Y'^{1:N})^2)} + \sum_{i\in\mathcal{S}_1} \sqrt{2\ln 2(Z(U_1^i|U_1^{1:i-1}) - Z(U_1^i|U_1^{1:i-1},Y^{1:N})^2)} \\
&\overset{(g)}{\leq} 2N\sqrt{4\ln 2\cdot 2^{-N^\beta}} = O(2^{-N^{\beta'}}),
\end{aligned}
$$

where $D(\cdot||\cdot)$ is the relative entropy, and the equalities and the inequalities follow from

(a) The telescoping expansion [30].

(b) $Q'(u_1^i|u_1^{1:i-1},y'^{1:N}) = P(u_1^i|u_1^{1:i-1},y'^{1:N})$ for $i \in \mathcal{I}_1$.

(c) Pinsker's inequality.

(d) Jensen's inequality.

(e) $Q'(u_1^i|u_1^{1:i-1}) = \frac{1}{2}$ ($\bar{u}_1^i$ is uniformly random) for $i \in \mathcal{F}_1$ and $Q'_{U_1^i|U_1^{1:i-1},Y'^{1:N}} = P_{U_1^i|U_1^{1:i-1}}$ for $i \in \mathcal{S}_1$.

(f) $Z(X|Y)^2 < H(X|Y) < Z(X|Y)$.

(g) (12).

Following the same fashion,

$$\begin{aligned}
&2\mathbb{V}(Q'_{U_1^{1:N},Y'^{1:N}}, Q_{U_1^{1:N},Y'^{1:N}}) \\
&\leq \sum_{i\in\mathcal{S}_1} \sqrt{2\ln 2 D(Q_{U_1^i}||Q'_{U^i}|U_1^{1:i-1},Y'^{1:N})} \\
&\overset{(h)}{=} \sum_{i\in\mathcal{S}_1} \sqrt{2\ln 2 (H(U_1^i|U_1^{1:i-1}) - 0)} \\
&\leq \sum_{i\in\mathcal{S}_1} \sqrt{2\ln 2 Z(U_1^i|U_1^{1:i-1})} \leq N\sqrt{2\ln 2 \cdot 2^{-N^\beta}} = O(2^{-N^{\beta'}}),
\end{aligned}$$

where inequality (h) follows from the MAP decision in (14) for $i \in \mathcal{S}_1$.

Finally, we have

$$\begin{aligned}
&\mathbb{V}(P_{U_1^{1:N},Y'^{1:N}}, Q_{U_1^{1:N},Y'^{1:N}}) \\
&\leq \mathbb{V}(P_{U_1^{1:N},Y'^{1:N}}, Q'_{U_1^{1:N},Y'^{1:N}}) + \mathbb{V}(Q'_{U_1^{1:N},Y'^{1:N}}, Q_{U_1^{1:N},Y'^{1:N}}) \\
&= O(2^{-N^{\beta'}}).
\end{aligned}$$

Clearly, when N goes to infinity, for any $R > \frac{|\mathcal{I}_1|}{N} = I(X_1;Y')$, $\mathbb{V}(P_{U_1^{1:N},Y'^{1:N}}, Q_{U_1^{1:N},Y'^{1:N}})$ is arbitrarily small. □

APPENDIX B
PROOF OF THEOREM 3

Proof: The variational distance can be upper bounded as follows.

$$\begin{aligned}
&2\mathbb{V}(P_{U_2^{1:N},U_1^{1:N},Y'^{1:N}}, Q_{U_2^{1:N},U_1^{1:N},Y'^{1:N}}) \\
&= \sum_{u_2^{1:N},u_1^{1:N},y'^{1:N}} |Q(u_2^{1:N},u_1^{1:N},y'^{1:N}) - P(u_2^{1:N},u_1^{1:N},y'^{1:N})| \\
&= \sum_{u_2^{1:N},u_1^{1:N},y'^{1:N}} |P(u_2^{1:N}|u_1^{1:N},y'^{1:N})P(u_1^{1:N},y'^{1:N}) - Q(u_2^{1:N}|u_1^{1:N},y'^{1:N})Q(u_1^{1:N},y'^{1:N})| \\
&\leq \sum_{u_2^{1:N},u_1^{1:N},y'^{1:N}} |P(u_2^{1:N}|u_1^{1:N},y'^{1:N}) - Q(u_2^{1:N}|u_1^{1:N},y'^{1:N})|P(u_1^{1:N},y'^{1:N}) \\
&\qquad + \sum_{u_2^{1:N},u_1^{1:N},y'^{1:N}} |P(u_1^{1:N},y'^{1:N}) - Q(u_1^{1:N},y'^{1:N})|Q(u_2^{1:N}|u_1^{1:N},y'^{1:N}).
\end{aligned} \tag{65}$$

Treating $(U_1^{1:N}, Y'^{1:N})$ as a new source with distribution $P(u_1^{1:N}, y'^{1:N})$, the first summation can be proved to be $O(2^{-N^{\beta'}})$ in the same fashion as the proof of Theorem 2. For the second summation, we have

$$\begin{aligned}
&\sum_{u_{1:2}^{1:N},y'^{1:N}} |P(u_1^{1:N},y'^{1:N}) - Q(u_1^{1:N},y'^{1:N})|Q(u_2^{1:N}|u_1^{1:N},y'^{1:N}) \\
&= \sum_{u_1^{1:N},y'^{1:N}} |P(u_1^{1:N},y'^{1:N}) - Q(u_1^{1:N},y'^{1:N})| \\
&= 2\mathbb{V}(P_{U_1^{1:N},Y'^{1:N}}, Q_{U_1^{1:N},Y'^{1:N}}) = O(2^{-N^{\beta'}}).
\end{aligned} \tag{66}$$

Finally,

$$\mathbb{V}(P_{U_2^{1:N},U_1^{1:N},Y'^{1:N}}, Q_{U_2^{1:N},U_1^{1:N},Y'^{1:N}}) = O(2 \cdot 2^{-N^{\beta'}}).$$

□

APPENDIX C
PROOF OF THEOREM 4

Proof: Firstly, for the source Y', we consider the average performance of the multilevel polar codes with all possible choice of $u_\ell^{\mathcal{F}_\ell}$ at each level. If the encoding rule described in the form of (17) is used for all $i \in [N]$ at each level, the resulted average distortion is given by

$$D_{P,Y'} = \frac{1}{N} \sum_{u_{1:r}^{1:N}, y'^{1:N}} P_{U_{1:r}^{1:N},Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N}) d(y'^{1:N}, \mathcal{M}(u_{1:r}^{1:N} G_N)),$$

where $\mathcal{M}(u_{1:r}^{1:N} G_N)$ denotes a mapping from $u_{1:r}^{1:N}$ to $x^{1:N}$ according to (38) (remind that x is drawn from Λ according to D_{Λ,σ_r}). For instance, let $\Lambda = \mathbb{Z}$ and the partition is given by $\mathbb{Z}/2\mathbb{Z}/...2^r\mathbb{Z}$, then $x^{1:N} = x_1^{1:N} + 2x_2^{1:N} + ... + 2^{r-1}x_r^{1:N}$ and $x_\ell^{1:N} = u_\ell^{1:N} G_N$. When $r \to \infty$, there exists a one-to-one mapping from $u_{1:r}^{1:N}$ to $x^{1:N}$. Then we have

$$\begin{aligned}
D_{P,Y'} &= \frac{1}{N} \sum_{x^{1:N}, y'^{1:N}} P_{X^{1:N},Y'^{1:N}}(x^{1:N}, y'^{1:N}) d(y'^{1:N}, x^{1:N}) \\
&= \frac{1}{N} \cdot N \sum_{x,y'} P_{X,Y'}(x,y') d(x,y') \\
&= \sum_{x \in \Lambda} P_X(x) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\Delta}} \exp(-\frac{(y'-x)^2}{2\Delta})(y'-x)^2 dy' \\
&= \Delta.
\end{aligned}$$

The result $D_{P,Y'} = \Delta$ is reasonable since the encoder does not do any compression. If we replace $P_{U_{1:r}^{1:N},Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N})$ with $Q_{U_{1:r}^{1:N},Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N})$ and compress $y'^{1:N}$ to $u_\ell^{\mathcal{I}_\ell}$ at each level, the resulted average distortion $D_{Q,Y'}$ can be bounded as

$$\begin{aligned}
D_{Q,Y'} &= \frac{1}{N} \sum_{u_{1:r}^{1:N}, y'^{1:N}} Q_{U_{1:r}^{1:N},Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N}) d(y'^{1:N}, \mathcal{M}(u_{1:r}^{1:N} G_N)) \\
&\leq \frac{1}{N} \sum_{u_{1:r}^{1:N}, y'^{1:N}} P_{U_{1:r}^{1:N},Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N}) d(y'^{1:N}, \mathcal{M}(u_{1:r}^{1:N} G_N)) \\
&\quad + \frac{1}{N} \sum_{u_{1:r}^{1:N}, y'^{1:N}} |P_{U_{1:r}^{1:N},Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N}) - Q_{U_{1:r}^{1:N},Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N})| d(y'^{1:N}, \mathcal{M}(u_{1:r}^{1:N} G_N))
\end{aligned}$$

Since the densities of both Y' and X decrease exponentially to their square norms, the distortion caused by large x or y' is negligible, we can always assume a maximum distortion $d_{\max}$ between y' and x. Then we have

$$\begin{aligned}
D_{Q,Y'} &\leq D_{P,Y'} + \frac{2}{N} \mathbb{V}(P_{U_{1:r}^{1:N},Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N}), Q_{U_{1:r}^{1:N},Y'^{1:N}}(u_{1:r}^{1:N}, y'^{1:N})) \cdot N d_{\max} \\
&= \Delta + O(2^{-N^{\beta'}}),
\end{aligned}$$

where the equation follows from (20) and $r = O(\log\log N)$ [11, Lemma 5].

Now we consider using the same encoder to quantize the Gaussian source Y. The resulted average distortion $D_{Q,Y}$ can be written as

$$\begin{aligned} D_{Q,Y} &= \frac{1}{N} \sum_{u_{1:r}^{1:N}, y^{1:N}} Q_{U_{1:r}^{1:N}, Y^{1:N}}(u_{1:r}^{1:N}, y^{1:N}) d(y^{1:N}, \mathcal{M}(u_{1:r}^{1:N} G_N)) \\ &= \frac{1}{N} \sum_{u_{1:r}^{1:N}, y^{1:N}} P_{Y^{1:N}}(y^{1:N}) Q_{U_{1:r}^{1:N}|Y^{1:N}}(u_{1:r}^{1:N}|y^{1:N}) \cdot d(y^{1:N}, \mathcal{M}(u_{1:r}^{1:N} G_N)). \end{aligned}$$

Since the same encoder is used, for a same realization $y^{1:N}$, we have $Q_{U_{1:r}^{1:N}|Y^{1:N}}(u_{1:r}^{1:N}|y^{1:N}) = Q_{U_{1:r}^{1:N}|Y'^{1:N}}(u_{1:r}^{1:N}|y^{1:N})$, and hence

$$\begin{aligned} D_{Q,Y} - D_{Q,Y'} &= \frac{1}{N} \sum_{u_{1:r}^{1:N}, y^{1:N}} (P_{Y^{1:N}}(y^{1:N}) - P_{Y'^{1:N}}(y^{1:N})) \cdot Q_{U_{1:r}^{1:N}|Y^{1:N}}(u_{1:r}^{1:N}|y^{1:N}) d(y^{1:N}, \mathcal{M}(u_{1:r}^{1:N} G_N)) \\ &\le \frac{1}{N} N d_{\max} \sum_{y^{1:N}} |P_{Y^{1:N}}(y^{1:N}) - P_{Y'^{1:N}}(y^{1:N})|. \end{aligned}$$

Again, by the telescoping expansion,

$$\begin{aligned} &\sum_{y^{1:N}} |P_{Y^{1:N}}(y^{1:N}) - P_{Y'^{1:N}}(y^{1:N})| \\ &= \sum_{y^{1:N}} \sum_{i=1}^{N} |P_{Y^i}(y^i) - P_{Y'^i}(y^i)| P_{Y^{1:i-1}}(y^{1:i-1}) P_{Y'^{i+1:N}}(y^{i+1:N}) \\ &= \sum_{i=1}^{N} \sum_{y^i} |P_{Y^i}(y^i) - P_{Y'^i}(y^i)| \\ &\underset{Lemma\,1}{\le} N \cdot 8\epsilon_\Lambda(\tilde{\sigma}_\Delta). \end{aligned}$$

As a result,

$$D_{Q,Y} \le \Delta + O(2^{-N^{\beta'}}) + 8\epsilon_\Lambda(\tilde{\sigma}_\Delta) d_{\max} N. \tag{67}$$

By scaling Λ, we can make $\epsilon_\Lambda(\tilde{\sigma}_\Delta) \ll \frac{1}{8 d_{\max} N}$, and $D_{Q,Y}$ can be arbitrarily close to Δ with $R > I(X;Y') \ge \frac{1}{2}\log\frac{\sigma_s^2}{\Delta} - \frac{5\epsilon_\Lambda(\tilde{\sigma}_\Delta)}{n}$ (n could be 1). When $\epsilon_\Lambda(\tilde{\sigma}_\Delta) \to 0$, we have $I(X;Y') \to \frac{1}{2}\log\frac{\sigma_s^2}{\Delta}$ and $R > \frac{1}{2}\log\frac{\sigma_s^2}{\Delta}$.

Now it is ready to explain the lattice structure. From the definition of $\mathcal{F}_\ell$ and [11, Lemma 6], it is easy to find that $\mathcal{F}_\ell \subseteq \mathcal{F}_{\ell-1}$ for $1 < \ell \le r$. When $u_\ell^{\mathcal{S}_\ell}$ is uniformly selected and $u_\ell^{\mathcal{F}_\ell} = 0$ at each level, the constructed polar code at level $\ell - 1$ is a subset of the polar code at level ℓ. Therefore, the resulted multilevel code is actually a polar lattice and the MAP decision on the bits in $\mathcal{S}_\ell$ is a shaping operation according to D_{Λ,σ_r}. Moreover, since $D_{Q,Y}$ is an average distortion over all random choices of $u_\ell^{\mathcal{F}_\ell}$, there exists at least one specific choice of $u_\ell^{\mathcal{F}_\ell}$ at each level making the average distortion satisfying (67). This is exactly a shift on the constructed polar lattice. Consequently, the shifted polar lattice achieves the rate-distortion bound of the Gaussian source. □

APPENDIX D
PROOF OF THEOREM 5

Proof: We firstly show that the target distortion can be achieved. Recall $U_\ell^{1:N} = A_\ell^{1:N} G_N$ for each level ℓ. Let $P_{U_{1:r}^{1:N},\bar{X}^{1:N}}$ denote the joint distribution between $U_{1:r}^{1:N}$ and $\bar{X}_{1:r}^{1:N}$ when the encoder performs no compression at each level, i.e., the encoder applies encoding rule (13) for all indices $i \in [N]$ at level 1, encoding rule (17) for all $i \in [N]$ at level 2 and similar rules for higher levels, with the notation X and Y' being replaced by A and $\bar{X}$, respectively. Let $Q_{U_{1:r}^{1:N},\bar{X}^{1:N}}$ denote the joint distribution when only $U_\ell^{\mathcal{I}_\ell^Q}$ is recorded following the random rounding rule at each level. $U_\ell^{\mathcal{F}_\ell^Q}$ is a uniformly random sequence shared between the encoder and decoder, and $U_\ell^{\mathcal{S}_\ell^Q}$ is determined according to the MAP rule (see (14) and (18)). As illustrated in (20),

$$\mathbb{V}(P_{U_{1:r}^{1:N},\bar{X}^{1:N}}, Q_{U_{1:r}^{1:N},\bar{X}^{1:N}}) = O(r \cdot 2^{-N^{\beta'}}). \tag{68}$$

Since $r = O(\log N)$, we can write $\mathbb{V}(P_{U_{1:r}^{1:N},\bar{X}^{1:N}}, Q_{U_{1:r}^{1:N},\bar{X}^{1:N}}) = O(2^{-N^{\beta'}})$. When quantization is performed for the source X, let $Q_{U_{1:r}^{1:N},X^{1:N}}$ denote the resulted joint distribution. By Lemma 1 again,

$$\begin{aligned}
&\sum_{u_{1:r}^{1:N},x^{1:N}} |Q_{U_{1:r}^{1:N},\bar{X}^{1:N}}(u_{1:r}^{1:N}, x^{1:N}) - Q_{U_{1:r}^{1:N},X^{1:N}}(u_{1:r}^{1:N}, x^{1:N})| \\
&= \sum_{x^{1:N}} |P_{\bar{X}^{1:N}}(x^{1:N}) - P_{X^{1:N}}(x^{1:N})| \sum_{u_{1:r}^{1:N}} Q_{U_{1:r}^{1:N}|X^{1:N}}(u_{1:r}^{1:N}|x^{1:N}) \\
&= \sum_{x^{1:N}} |P_{\bar{X}^{1:N}}(x^{1:N}) - P_{X^{1:N}}(x^{1:N})| \leq N \cdot 8\epsilon_\Lambda(\tilde{\sigma}_q).
\end{aligned}$$

It has been shown that $\epsilon_\Lambda(\tilde{\sigma}_q) = O(2^{-\sqrt{N}})$, we further have

$$\begin{aligned}
\mathbb{V}(P_{U_{1:r}^{1:N},\bar{X}^{1:N}}, Q_{U_{1:r}^{1:N},X^{1:N}}) &\leq \mathbb{V}(P_{U_{1:r}^{1:N},\bar{X}^{1:N}}, Q_{U_{1:r}^{1:N},\bar{X}^{1:N}}) + \mathbb{V}(Q_{U_{1:r}^{1:N},\bar{X}^{1:N}}, Q_{U_{1:r}^{1:N},X^{1:N}}) &(69)\\
&= O(2^{-N^{\beta'}}) + O(2^{-\sqrt{N}}) = O(2^{-N^{\beta'}}). &(70)
\end{aligned}$$

As mentioned in Section V, the encoder only sends $U_\ell^{d\mathcal{I}_\ell}$ to the decoder, which then utilizes the side information to recover $U_\ell^{\mathcal{I}_\ell^C}$. Here we assume that $U_\ell^{\mathcal{I}_\ell^C}$ can be correctly decoded and $U_\ell^{\mathcal{S}_\ell^Q}$ is recovered according to the MAP rule. In this case, the decoder enjoys the same joint distribution $Q_{U_{1:r}^{1:N},X^{1:N}}$ as the encoder does. Recall that $B = \frac{\alpha_q}{\alpha_c} Y$ and $\bar{B} = \frac{\alpha_q}{\alpha_c} \bar{Y}$. Let $Q_{U_{1:r}^{1:N},X^{1:N},B^{1:N}}$ denote the resulted joint distribution of $U_{1:r}^{1:N}$, $X^{1:N}$, and $B^{1:N}$ when the encoder performs compression, i.e., compresses $X^{1:N}$ to $U_\ell^{\mathcal{I}_\ell^Q}$ at each level. Let $P_{U_{1:r}^{1:N},\bar{X}^{1:N},\bar{B}^{1:N}}$ denote the resulted joint distribution of $U_{1:r}^{1:N}$, $\bar{X}^{1:N}$, and $\bar{B}^{1:N}$ when the encoder performs no compression for $\bar{X}^{1:N}$.

$$\begin{aligned}
&2\mathbb{V}(P_{U_{1:r}^{1:N},\bar{X}^{1:N},\bar{B}^{1:N}}, Q_{U_{1:r}^{1:N},X^{1:N},B^{1:N}}) \\
&= \sum_{u_{1:r}^{1:N},x^{1:N},b^{1:N}} |P(u_{1:r}^{1:N}, x_{1:r}^{1:N}, b^{1:N}) - Q(u_{1:r}^{1:N}, x_{1:r}^{1:N}, b^{1:N})| \\
&= \sum_{u_{1:r}^{1:N},x^{1:N},b^{1:N}} |P(u_{1:r}^{1:N}, x_{1:r}^{1:N}) P(b^{1:N}|u_{1:r}^{1:N}, x_{1:r}^{1:N}) - Q(u_{1:r}^{1:N}, x_{1:r}^{1:N}) Q(b^{1:N}|u_{1:r}^{1:N}, x_{1:r}^{1:N})|.
\end{aligned} \tag{71}$$

According to Fig. 7, $\alpha_q X' \to X \to B$ and $A \to \bar{X} \to \bar{B}$ are two Markov chains. We have

$$\begin{aligned} P(b^{1:N}|u_{1:r}^{1:N}, x_{1:r}^{1:N}) &= Q(b^{1:N}|u_{1:r}^{1:N}, x_{1:r}^{1:N}) && (72)\\ &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_{z'}^2}} \exp(-\frac{(b^i - x^i)^2}{2\sigma_{z'}^2}). && (73)\end{aligned}$$

Therefore,

$$\mathbb{V}(P_{U_{1:r}^{1:N}, \bar{X}^{1:N}, \bar{B}^{1:N}}, Q_{U_{1:r}^{1:N}, X^{1:N}, B^{1:N}}) = \mathbb{V}(P_{U_{1:r}^{1:N}, \bar{X}^{1:N}}, Q_{U_{1:r}^{1:N}, X^{1:N}}) = O(2^{-N^{\beta'}}). \quad (74)$$

Recall that the reconstruction of $\bar{X}$ is given by $\check{X} = A + \gamma(\bar{B} - A)$. The average distortion Δ_P caused by $P_{U_{1:r}^{1:N}, \bar{X}^{1:N}, \bar{B}^{1:N}}$ can be expressed as

$$\Delta_P = \frac{1}{N} \sum_{u_{1:r}^{1:N}, x^{1:N}, b^{1:N}} P_{U_{1:r}^{1:N}, \bar{X}^{1:N}, \bar{B}^{1:N}}(u_{1:r}^{1:N}, x^{1:N}, b^{1:N}) d(x^{1:N}, \check{x}^{1:N}), \quad (75)$$

where $\check{x}^{1:N} = \gamma b^{1:N} + (1-\gamma)\mathcal{M}(u_{1:r}^{1:N} G_N)$, where $\mathcal{M}(u_{1:r}^{1:N} G_N)$ is a mapping from $u_{1:r}^{1:N}$ to $a^{1:N}$ according to the lattice Gaussian distribution. Clearly, given $u_{1:r}^{1:N}$, there is a one-to-one mapping between $b^{1:N}$ and $\check{x}^{1:N}$ when r is sufficiently large. Thus, Δ_P can be written as

$$\begin{aligned} \Delta_P &= \frac{1}{N} \sum_{u_{1:r}^{1:N}, \check{x}^{1:N}, x^{1:N}} P_{U_{1:r}^{1:N}, \check{X}^{1:N}, \bar{X}^{1:N}}(u_{1:r}^{1:N}, \check{x}^{1:N}, x^{1:N},) d(\check{x}^{1:N}, x^{1:N})\\ &= \frac{1}{N} \sum_{\check{x}^{1:N}, x^{1:N}} P_{\check{X}^{1:N}, \bar{X}^{1:N}}(\check{x}^{1:N}, x^{1:N}) d(\check{x}^{1:N}, x^{1:N})\\ &= \frac{1}{N} \cdot N \sum_{\check{x}, x} P_{\check{X}, \bar{X}}(\check{x}, x) d(\check{x}, x)\\ &= \int_{\check{x}} f_{\check{X}}(\check{x}) \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\Delta}} \exp(-\frac{(x-\check{x})^2}{2\Delta})(x-\check{x})^2 dx d\check{x}\\ &= \Delta. \end{aligned}$$

The expected distortion Δ_Q achieved by $Q_{U_{1:r}^{1:N}, X^{1:N}, B^{1:N}}$ satisfies

$$\begin{aligned} \Delta_Q - \Delta_P &= \frac{1}{N} \sum_{u_{1:r}^{1:N}, x^{1:N}, b^{1:N}} (P_{U_{1:r}^{1:N}, \bar{X}^{1:N}, \bar{B}^{1:N}} - Q_{U_{1:r}^{1:N}, X^{1:N}, B^{1:N}}) d(\check{x}^{1:N}, x^{1:N})\\ &\le \frac{1}{N} N d_{\max} \sum_{y^{1:N}} |P_{U_{1:r}^{1:N}, \bar{X}^{1:N}, \bar{B}^{1:N}} - Q_{U_{1:r}^{1:N}, X^{1:N}, B^{1:N}}|\\ &= O(2^{-N^{\beta'}}). \end{aligned}$$

Now we show that the decoder is able to decode $U_\ell^{\mathcal{I}_\ell^C}$ with vanishing error probability.

$$\begin{aligned} 2\mathbb{V}(P_{U_{1:r}^{1:N}, \bar{B}^{1:N}}, Q_{U_{1:r}^{1:N}, B^{1:N}}) &= \sum_{u_{1:r}^{1:N}, b^{1:N}} |P(u_{1:r}^{1:N}, b^{1:N}) - Q(u_{1:r}^{1:N}, b^{1:N})|\\ &= \sum_{u_{1:r}^{1:N}, b^{1:N}} |\sum_{x^{1:N}} \left[P(u_{1:r}^{1:N}, x^{1:N}, b^{1:N}) - Q(u_{1:r}^{1:N}, x^{1:N}, b^{1:N})\right]|\\ &\le \sum_{u_{1:r}^{1:N}, b^{1:N}} \sum_{x^{1:N}} |P_{U_{1:r}^{1:N}, \bar{X}^{1:N}, \bar{B}^{1:N}} - Q_{U_{1:r}^{1:N}, X^{1:N}, B^{1:N}}|\\ &= O(2^{-N^{\beta'}}). \quad (76)\end{aligned}$$

By the result of [11], $P_{U_{1:r}^{1:N},\bar{B}^{1:N}}$ results in an expectation of error probability $E_P^\ell[P_e]$ at each level such that $E_P^\ell[P_e] = O(2^{-N^{\beta'}})$. To see this, let $\mathcal{E}_i$ denote the set of pairs of $u_\ell^{1:N}$ and $b^{1:N}$ such that the SC decoding error occurs at the ith bit for level ℓ, then the block decoding error event is given by $\mathcal{E}^\ell \equiv \bigcup_{i\in\mathcal{I}_\ell^C} \mathcal{E}_i$. Then the expectation of decoding error probability over all random mapping is expressed as

$$\begin{aligned}
E_P^\ell[P_e] &= \sum_{u_{1:\ell}^{1:N},b^{1:N}} P_{U_{1:\ell}^{1:N},\bar{B}^{1:N}}(u_{1:\ell}^{1:N},b^{1:N})\mathbb{1}[(u_\ell^{1:N},b^{1:N})\in\mathcal{E}^\ell] \\
&\le \sum_{i\in\mathcal{I}_\ell^C\cup\mathcal{S}_\ell^Q}\ \sum_{u_{1:\ell}^{1:N},b^{1:N}} P_{U_{1:\ell}^{1:N},\bar{B}^{1:N}}(u_{1:\ell}^{1:N},b^{1:N})\mathbb{1}[(u_\ell^{1:N},b^{1:N})\in\mathcal{E}_i] \\
&\le \sum_{i\in\mathcal{I}_\ell^C\cup\mathcal{S}_\ell^Q}\ \sum_{u_\ell^{1:i},u_{1:\ell-1}^{1:N},b^{1:N}} P(u_\ell^{1:i-1},u_{1:\ell-1}^{1:N},b^{1:N})P(u_\ell^i|u_\ell^{1:i-1},u_{1:\ell-1}^{1:N},b^{1:N}) \\
&\qquad\qquad \cdot\mathbb{1}[P(u_\ell^i|u_\ell^{1:i-1},u_{1:\ell-1}^{1:N},b^{1:N}) \le P(u_\ell^i\oplus 1|u_\ell^{1:i-1},u_{1:\ell-1}^{1:N},b^{1:N})] \\
&\le \sum_{i\in\mathcal{I}_\ell^C\cup\mathcal{S}_\ell^Q}\ \sum_{u_\ell^{1:i},u_{1:\ell-1}^{1:N},b^{1:N}} P(u_\ell^{1:i-1},u_{1:\ell-1}^{1:N},b^{1:N})P(u_\ell^i|u_\ell^{1:i-1},u_{1:\ell-1}^{1:N},b^{1:N}) \\
&\qquad\qquad \sqrt{\frac{P(u_\ell^i\oplus 1|u_\ell^{1:i-1},u_{1:\ell-1}^{1:N},b^{1:N})}{P(u_\ell^i|u_\ell^{1:i-1},u_{1:\ell-1}^{1:N},b^{1:N})}} \\
&\le N\cdot Z(U_\ell^i|U_\ell^{1:i-1},A_{1:\ell-1}^{1:N},\bar{B}^{1:N}) \\
&= O(2^{-N^{\beta'}}).
\end{aligned} \tag{77}$$

Then by the union bound, we immediately obtain that the expectation of multistage decoding error probability for the polar lattice $E_P[P_e] = O(2^{-N^{\beta'}})$. Let P_e^{WZ} denote the expectation of error probability caused by $Q_{U_{1:r}^{1:N},B^{1:N}}$, i.e., it is an average error probability over all choices of the frozen bits $U_\ell^{\mathcal{F}_\ell^C}$ and shaping bits $U_\ell^{d\mathcal{S}_\ell}$ for each level. Let $\mathcal{E}$ denote the set of the pairs $(u_{1:r}^{1:N},b^{1:N})$ such that a lattice decoding error occurs. We have

$$\begin{aligned}
P_e^{WZ} - E_P[P_e] &= \sum_{u_{1:\ell}^{1:N},b^{1:N}} (P(u_{1:\ell}^{1:N},b^{1:N}) - Q(u_{1:\ell}^{1:N},b^{1:N}))\cdot\mathbb{1}[(u_{1:r}^{1:N},b^{1:N})\in\mathcal{E}] && (78)\\
&\le 2\mathbb{V}(P_{U_{1:r}^{1:N},\bar{B}^{1:N}}, Q_{U_{1:r}^{1:N},B^{1:N}}) && (79)\\
&\le O(2^{-N^{\beta'}}). && (80)
\end{aligned}$$

With regard to the data rate, we have

$$\sum_{\ell=1}^r \frac{|\mathcal{I}_\ell^Q|}{N} \to \frac{1}{2}\log\left(\frac{\sigma_x^2\sigma_z^2-\sigma_y^2\Delta}{\sigma_z^2\Delta}\right)^+, \tag{81}$$

and

$$\sum_{\ell=1}^r \frac{|\mathcal{I}_\ell^C|}{N} \to \frac{1}{2}\log\left(\frac{\sigma_x^2\sigma_z^2-\sigma_y^2\Delta}{\sigma_z^4}\right)^-. \tag{82}$$

Finally,

$$R = \sum_{\ell=1}^r \frac{|\mathcal{I}_\ell^Q|}{N} - \frac{|\mathcal{I}_\ell^C|}{N} \to \frac{1}{2}\log(\frac{\sigma_z^2}{\Delta})^+. \tag{83}$$

□

APPENDIX E
PROOF OF THEOREM 6

Proof: We firstly show the power constraint P can be satisfied. Recall $U_\ell^{1:N} = A_\ell^{1:N} G_N$ for each level ℓ. Similar to the previous proof, denote by $P_{U_{1:r}^{1:N}, \bar{T}^{1:N}}$ the joint distribution between $U_{1:r}^{1:N}$ and $\bar{T}_{1:r}^{1:N}$ when the encoder applies random rounding rule for all indices $i \in [N]$ at level ℓ. Denote by $Q_{U_{1:r}^{1:N}, T^{1:N}}$ the joint distribution when $U_\ell^{\mathcal{I}_\ell^Q}$ and $U_\ell^{d\mathcal{S}_\ell}$ are encoded following the random rounding rule at each level. $U_\ell^{\mathcal{F}_\ell^C}$ is a uniformly random sequence shared between the encoder and decoder, $U_\ell^{d\mathcal{F}_\ell}$ is a uniform message sequence and $U_\ell^{\mathcal{S}_\ell^C}$ is determined according to the MAP rule.

Notice that $\epsilon_\Lambda(\tilde{\sigma}_q) \leq \epsilon_\Lambda(\tilde{\sigma}_c) = O(2^{-\sqrt{N}})$. Similar to the previous proof, we have

$$\mathbb{V}(P_{U_{1:r}^{1:N}, \bar{T}^{1:N}}, Q_{U_{1:r}^{1:N}, T^{1:N}}) \leq O(2^{-N^{\beta'}}) + O(2^{-\sqrt{N}}) = O(2^{-N^{\beta'}}). \tag{84}$$

Thus, the average transmit power realized by $Q_{U_{1:r}^{1:N}, T^{1:N}}$ can be arbitrarily close to that realized by $P_{U_{1:r}^{1:N}, \bar{T}^{1:N}}$, i.e.,

$$\frac{1}{N} \sum_{u_{1:r}^{1:N}, t^{1:N}} [Q_{U_{1:r}^{1:N}, T^{1:N}}(u_{1:r}^{1:N}, t^{1:N}) - P_{U_{1:r}^{1:N}, \bar{T}^{1:N}}(u_{1:r}^{1:N}, t^{1:N})] d(\frac{1}{\alpha_q} \mathcal{M}(u_{1:r}^{1:N} G_N), t^{1:N}) \tag{85}$$

$$\leq \frac{2}{N} N d_{max} \mathbb{V}(P_{U_{1:r}^{1:N}, \bar{T}^{1:N}}, Q_{U_{1:r}^{1:N}, T^{1:N}}), \tag{86}$$

$$\leq O(2^{-N^{\beta'}}), \tag{87}$$

where $\mathcal{M}(u_{1:r}^{1:N} G_N)$ denotes a mapping from $u_{1:r}^{1:N}$ to $t^{1:N}$ according to the lattice Gaussian distribution. Note that for a constant σ_a^2 and σ_s^2, the assumption of a maximum distortion d_{max} is reasonable since the Gaussian distribution decays exponentially with large distortion.

Now we show that the average transmit power realized by $P_{U_{1:r}^{1:N}, \bar{T}^{1:N}}$ is arbitrarily close to P. When r is sufficiently large, $P_{A_{1:r}} \to D_{\Lambda, \sigma_a^2}$, and $\bar{T} = A + N(0, \alpha_q P)$ as shown in Fig. 11. Then the variable $\frac{1-\alpha_q}{\alpha_q} A + (A - \bar{T}) = \frac{1-\alpha_q}{\alpha_q} A + N(0, \alpha_q P)$ corresponds to a variable resulted from adding a lattice Gaussian distributed variable to an independent Gaussian noise. Notice that $\frac{1-\alpha_q}{\alpha_q}$ only involves a scale on D_{Λ, σ_a^2}. When $\epsilon_\Lambda(\tilde{\sigma}_q) \leq O(2^{-\sqrt{N}})$, the flatness factor associated with the AWGN channel from $\frac{1-\alpha_q}{\alpha_q} A$ to $\frac{1-\alpha_q}{\alpha_q} A + N(0, \alpha_q P)$ is also upper bounded by $O(2^{-\sqrt{N}})$.

Check that $\left(\frac{1-\alpha_q}{\alpha_q}\right)^2 \sigma_a^2 = (1 - \alpha_q) P$. Let $\dot{X}$ and $\ddot{X}$ denote $\frac{1-\alpha_q}{\alpha_q} A + N(0, \alpha_q P)$ and Gaussian random variable with distribution $N(0, P)$, respectively. By Lemma 1, $\mathbb{V}(P_{\dot{X}}, P_{\ddot{X}}) \leq O(2^{-\sqrt{N}})$. Let $x^{1:N} = \frac{1}{\alpha_q} \mathcal{M}(u_{1:r}^{1:N} G_N) - t^{1:N}$, we have

$$\frac{1}{N} \sum_{u_{1:r}^{1:N}, t^{1:N}} P_{U_{1:r}^{1:N}, \bar{T}^{1:N}}(u_{1:r}^{1:N}, t^{1:N}) d(\frac{1}{\alpha_q} u_{1:r}^{1:N} G_N, t^{1:N}) \tag{88}$$

$$= \frac{1}{N} \sum_{x^{1:N}} P_{\dot{X}^{1:N}} d(x^{1:N}, 0) \tag{89}$$

$$= E_{\dot{X}}[x^2]. \tag{90}$$

Since $\mathbb{V}(P_{\dot{X}}, P_{\ddot{X}}) \leq O(2^{-\sqrt{N}})$, $E_{\dot{X}}[x^2] - E_{\ddot{X}}[x^2] \leq O(2^{-\sqrt{N}})$. Consequently,

$$\begin{aligned} P_T &= \frac{1}{N} \sum_{u_{1:r}^{1:N}, t^{1:N}} Q_{U_{1:r}^{1:N}, T^{1:N}}(u_{1:r}^{1:N}, t^{1:N}) d(\frac{1}{\alpha_q} \mathcal{M}(u_{1:r}^{1:N} G_N), t^{1:N}) && (91) \\ &\leq \frac{1}{N} \sum_{u_{1:r}^{1:N}, t^{1:N}} P_{U_{1:r}^{1:N}, \bar{T}^{1:N}}(u_{1:r}^{1:N}, t^{1:N}) d(\frac{1}{\alpha_q} \mathcal{M}(u_{1:r}^{1:N} G_N), t^{1:N}) + O(2^{-N^{\beta'}}) && (92) \\ &\leq E_{\ddot{X}}[x^2] + O(2^{-\sqrt{N}}) + O(2^{-N^{\beta'}}) && (93) \\ &= P + O(2^{-N^{\beta'}}). && (94) \end{aligned}$$

Now we prove the reliability. Recall that $Y = S + X + Z$, where $X = S' - \rho S$ is independent of S. Scaling Y by ρ gives us

$$\begin{aligned} \rho Y &= \rho S' + \rho(1-\rho)S + \rho Z && (95) \\ &= \alpha_c S' + \rho(1-\rho)S - (\alpha_c - \rho)S' + \rho Z. && (96) \end{aligned}$$

Check that

$$\alpha_c - \alpha_q = \frac{P(P + \sigma_z^2)}{P\sigma_i^2 + (P + \sigma_z^2)^2} = (1 - \alpha_q)\rho, \tag{97}$$

leaving us $\alpha_c - \rho = \alpha_q(1 - \rho)$. Scale ρY by $\frac{\alpha_q}{\alpha_c}$. Then

$$\frac{\alpha_q}{\alpha_c}\rho Y = \alpha_q S' + \frac{\alpha_q}{\alpha_c}(1-\rho)(\rho S - \alpha_q S') + \frac{\alpha_q}{\alpha_c}\rho Z. \tag{98}$$

Note that both $\rho S - \alpha_q S'$ and Z are independent of S'. Replacing $\alpha_q S'$ with A, we have

$$\frac{\alpha_q}{\alpha_c}\rho \dot{Y} = A + \frac{\alpha_q}{\alpha_c}(1-\rho)(\rho S - A) + \frac{\alpha_q}{\alpha_c}\rho Z, \tag{99}$$

which corresponds to the reverse solution shown in Fig. 11. Let $\bar{Y}$ denote the channel output when S is replaced by $\bar{S}$, i.e.,

$$\frac{\alpha_q}{\alpha_c}\rho \bar{Y} = A + \frac{\alpha_q}{\alpha_c}(1-\rho)(\rho \bar{S} - A) + \frac{\alpha_q}{\alpha_c}\rho Z. \tag{100}$$

Recall $\bar{T} = \rho\bar{S}$, $T = \rho S$, and $\bar{B} = \frac{\alpha_q}{\alpha_c}\rho\bar{Y}$. Also let $\dot{B} = \frac{\alpha_q}{\alpha_c}\rho\dot{Y}$. According to the previous proof, we already have $\mathbb{V}(P_{U_{1:r}^{1:N}, \bar{T}^{1:N}}, Q_{U_{1:r}^{1:N}, T^{1:N}}) \leq O(2^{-N^{\beta'}})$. Note that Z is an independent Gaussian noise, it is not difficult to obtain that

$$\mathbb{V}(P_{U_{1:r}^{1:N}, \bar{T}^{1:N}, \bar{B}^{1:N}}, Q_{U_{1:r}^{1:N}, T^{1:N}, \dot{B}^{1:N}}) = \mathbb{V}(P_{U_{1:r}^{1:N}, \bar{T}^{1:N}}, Q_{U_{1:r}^{1:N}, T^{1:N}}) \leq O(2^{-N^{\beta'}}), \tag{101}$$

since $P_{\bar{B}^{1:N}|U_{1:r}^{1:N}, \bar{T}^{1:N}} = Q_{\dot{B}^{1:N}|U_{1:r}^{1:N}, T^{1:N}}$. By (76),

$$\mathbb{V}(P_{U_{1:r}^{1:N}, \bar{B}^{1:N}}, Q_{U_{1:r}^{1:N}, \dot{B}^{1:N}}) \leq \mathbb{V}(P_{U_{1:r}^{1:N}, \bar{T}^{1:N}, \bar{B}^{1:N}}, Q_{U_{1:r}^{1:N}, T^{1:N}, \dot{B}^{1:N}}) \leq O(2^{-N^{\beta'}}). \tag{102}$$

Note that $\bar{B}$ is a variable obtained by adding A to a Gaussian noise, and $\dot{B}$ is the real signal received because the side information S is Gaussian distributed. Similar to (77), the expectation of error probability $E_P[P_e]$ caused

by $P_{U_{1:r}^{1:N},\bar{B}^{1:N}}$ can be unpper bounded as $E_P[P_e] \leq O(2^{-N^{\beta'}})$. Finally, the expectation of error probability P_e^{GP} caused by $Q_{U_{1:r}^{1:N},\dot{B}^{1:N}}$ satisfies

$$\begin{aligned} P_e^{GP} &\leq E_P[P_e] + 2\mathbb{V}(P_{U_{1:r}^{1:N},\bar{B}^{1:N}}, Q_{U_{1:r}^{1:N},\dot{B}^{1:N}}) && (103)\\ &\leq O(2^{-N^{\beta'}}). && (104)\end{aligned}$$

□

REFERENCES

[1] R. Gray, "Vector quantization," *IEEE ASSP Mag.*, vol. 1, no. 2, pp. 4–29, April 1984.

[2] J. H. Conway and N. J. A. Sloane, *Sphere Packings, Lattices, and Groups*. New York: Springer, 1993.

[3] B. Juang and J. Gray, A., "Multiple stage vector quantization for speech coding," in *IEEE Int. Conf. Acoustics, Speech, and Signal Process.*, vol. 7, May 1982, pp. 597–600.

[4] A. Gersho and V. Cuperman, "Vector quantization: A pattern-matching technique for speech coding," *IEEE Commun. Mag*, vol. 21, no. 9, pp. 15–21, Dec. 1983.

[5] M. Sabin and R. Gray, "Product code vector quantizers for waveform and voice coding," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 32, no. 3, pp. 474–488, June 1984.

[6] R. G. Gallager, *Information theory and reliable communication*. Springer, 1968, vol. 2.

[7] W. Finamore and W. Pearlman, "Optimal encoding of discrete-time continuous-amplitude memoryless sources with finite output alphabets," *IEEE Trans. Inf. Theory*, vol. 26, no. 2, pp. 144–155, Mar. 1980.

[8] M. Marcellin and T. Fischer, "Trellis coded quantization of memoryless and Gauss-markov sources," *IEEE Trans. Commun.*, vol. 38, no. 1, pp. 82–93, Jan. 1990.

[9] B. Kudryashov and K. Yurkov, "Near-optimum low-complexity lattice quantization," in *Proc. 2010 IEEE Int. Symp. Inform. Theory*, Austin, USA, June 2010, pp. 1032–1036.

[10] S. Vatedka and N. Kashyap, "Some "goodness" properties of LDA lattices," Oct. 2014. [Online]. Available: http://arxiv.org/abs/1410.7619

[11] Y. Yan, L. Liu, C. Ling, and X. Wu, "Construction of capacity-achieving lattice codes: Polar lattices," Nov. 2014. [Online]. Available: http://arxiv.org/abs/1411.0187

[12] C. Ling and L. Gan, "Lattice quantization noise revisited," in *Proc. 2013 IEEE Inform. Theory Workshop*, Sept. 2013, pp. 1–5.

[13] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1152–1159, July 1996.

[14] R. Zamir, *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory*. Cambridge, UK: Cambridge University Press, 2014.

[15] S. Korada and R. Urbanke, "Polar codes are optimal for lossy source coding," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1751–1768, April 2010.

[16] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan. 1976.

[17] R. Venkataramanan, T. Sarkar, and S. Tatikonda, "Lossy compression via sparse linear regression: Computationally efficient encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3265–3278, June 2014.

[18] R. Venkataramanan, A. Joseph, and S. Tatikonda, "Lossy compression via sparse linear regression: Performance under minimum-distance encoding," *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3254–3264, June 2014.

[19] R. Zamir, S. Shamai, and U. Erez, "Nested linear/lattice codes for structured multiterminal binning," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1250–1276, June 2002.

[20] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3820–3833, Nov. 2005.

[21] C. Ling, S. Gao, and J.-C. Belfiore, "Wyner-ziv coding based on multidimensional nested lattices," *IEEE Trans. Commun.*, vol. 60, no. 5, pp. 1328–1335, May 2012.

[22] U. Erez and S. ten Brink, “A close-to-capacity dirty paper coding scheme,” *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3417–3432, Oct. 2005.

[23] G. D. Forney Jr., M. Trott, and S.-Y. Chung, “Sphere-bound-achieving coset codes and multilevel coset codes,” *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 820–850, May 2000.

[24] C. Ling, L. Luzzi, J. Belfiore, and D. Stehle, “Semantically secure lattice codes for the Gaussian wiretap channel,” *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6399–6416, Oct. 2014.

[25] C. Ling and J. Belfiore, “Achieving AWGN channel capacity with lattice Gaussian coding,” *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 5918–5929, Oct. 2014.

[26] E. Arıkan, “Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels,” *IEEE Trans. Inf. Theory*, vol. 55, no. 7, pp. 3051–3073, July 2009.

[27] I. Tal and A. Vardy, “How to construct polar codes,” *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.

[28] R. Pedarsani, S. Hassani, I. Tal, and E. Telatar, “On the construction of polar codes,” in *Proc. 2011 IEEE Int. Symp. Inform. Theory (ISIT 2011)*, July 2011, pp. 11–15.

[29] E. Arıkan, “Source polarization,” in *Proc. 2010 IEEE Int. Symp. Inform. Theory*, Austin, USA, 2010, pp. 899–903.

[30] J. Honda and H. Yamamoto, “Polar coding without alphabet extension for asymmetric models,” *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 7829–7838, Dec. 2013.

[31] M. H. M. Costa, “Writing on dirty paper (corresp.),” *IEEE Trans. Inf. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.

[32] S. I. Gelfand and M. S. Pinsker, “Coding for channel with random parameters,” *Probl. Inform. Transm*, vol. 9, no. 1, pp. 19–31, 1980.